Spring 5-15-2014

A Study of Sampling Strategies for Helical CT

Daheng Li
Washington University in St. Louis

Washington University in St. Louis
School of Engineering and Applied Science
Department of Electrical and Systems Engineering

Thesis Examination Committee:
Joseph A. O'Sullivan
Hiroaki Mukai
Jr-Shin Li
David G. Politte

A Study of Sampling Strategies for Helical CT

by

Daheng Li

A thesis presented to the School of Engineering
of Washington University in partial fulfillment of the
requirements for the degree of

Master of Science

May 2014
Saint Louis, Missouri

ABSTRACT OF THE THESIS

A Study of Sampling Strategies for Helical CT

by

Daheng Li

Master of Science in Electrical Engineering

Washington University in St. Louis, 2013

Research Advisor: Professor Joseph O' Sullivan

Two classes of subsampling strategies, partially inspired by ideas from compressed sensing (CS), are developed and tested using real medical x-ray CT data acquired with a helical geometry. A version of the Feldkamp algorithm for helical x-ray CT is described. An alternating minimization (AM) algorithm for finding the maximum-likelihood estimates of attenuation functions in transmission X-ray tomography, developed by O'Sullivan and Benac, is then introduced. The derivation of this AM algorithm is extended to include an optional regularization term, which makes it a MAP estimate. A Newton's method with trust region modification is implemented for the regularization. In addition, the alternating minimization (AM) algorithm when using data from a subset of detectors, developed by Snyder, is illustrated. Ordered subsets techniques are used to increase the convergence rate. Results of subsampling strategies are demonstrated on real data by subsampling the actual measurements and reconstructing.

to my parents and my love Yihe

Contents

List of Figures

Acknowledgments

I would like to express the deepest appreciation to my research advisor Dr. O'Sullivan for his forward-looking ideas and significant advising. I would like to thank Dr. Politte for technical support and useful comments, remarks and engagement through the whole learning process of this master thesis. Furthermore, I would like to acknowledge Drs. O'Sullivan and Daniel Keesing for authorizing me to use the CT reconstruction C++ package HECTARE v.1.0 written by Dr. Keesing a few years ago and to use some figures from his publications. This package is significant since it is the starting point of most of the simulations in this thesis. I would like to thank Dr. Steven Don from St. Louis Children's Hospital and Washington University School of Medicine for acquiring the data for my experiments. I completely understand that giving data to engineers required lots of time, efforts and, most importantly, was very expensive. Also, I would like to thank my lab-mates Yaqi Chen, Soysal Degirmenci and Ikenna Odinaka for useful discussions on this topic. Also, Ikenna Odinaka wrote the C++ pseudo code for Newton's method with trust region, which was very helpful. In addition, a thank you goes to my academic advisor, Professor Martin Arthur, for course-selection suggestions and help. The courses are the fundamental and necessary knowledge of this thesis. Finally, I would like to thank Engineering IT for providing me the access to the Engineering Cloud Cluster, on which most of the computations in this thesis were done.

Daheng Li

Washington University in St. Louis
May 2014

Chapter 1

Introduction

Computed Tomography (CT) plays an important role in clinical diagnosis, and is also important in image-guided therapy [1][2]. Generally, a CT system is designed to measure some spatially-variant contrast, which is then used to distinguish materials. Today, important goals of diagnostic radiology are to minimize exposure of the patient to radiation and to minimize the time needed for data collection [3][4]. In general, for the same CT system, with more radiation dose (within a reasonable range) being used, higher image quality (higher signal-to-noise ratio, larger contrast, etc.) is obtainable. However, the scan will be more risky to the patients because of the additional radiation. The main risks are those associated with the increased possibility of cancer induction from x-ray radiation exposure. As CT instruments continue to evolve and novel geometries emerge, a need arises to understand the constraints on sampling of the data required to obtain images with diagnostically useful quality.

Currently available multi-detector-row CT scanners collect an abundance of data and can serve as test-beds to evaluate multiple sampling strategies. These strategies can be inspired by some ideas from compressed sensing (CS) theories [5][6][7][8][9][10]. Fewer measurements potentially will require lower dose and less time to acquire data, but the challenge is that when reconstructed with state-of-the-art image reconstruction algorithms, they must produce images of diagnostic quality [11].

The following three sections will focus on basic knowledge of the helical multi-detector-row CT, including the geometry, the data acquisition procedure and reconstructions.

1.1 Geometry

Helical multi-detector-row X-ray CT has proven to be a successful imaging modality and it is widely used in numerous clinical applications. Multiple rows of detectors are arranged in a cylindrical arc on the opposite side of the patient port from the x-ray source at its center. During scanning, the source and detector array rotate together as the patient is being transported through the scanner. Figure 1.1, plotted by Daniel Keesing, shows the basic CT geometry used in this thesis [12].

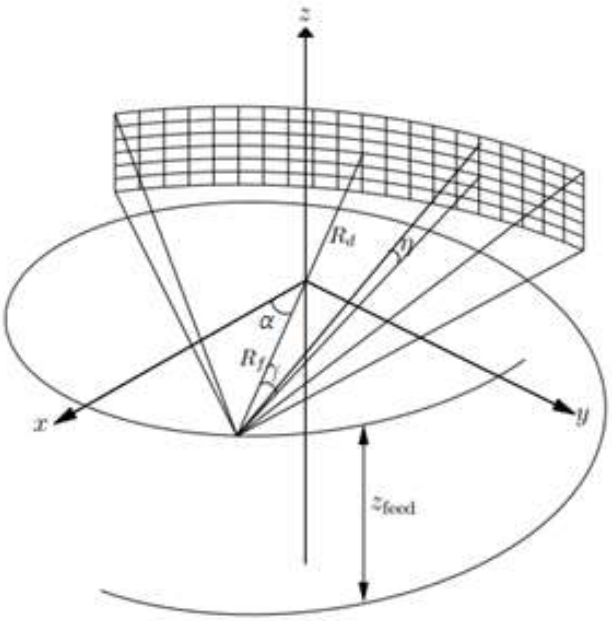


Figure 1.1 Geometry of the multi-detector-row X-ray CT used in this thesis.

α is the angle of the x-ray source with respect to the x axis. The fan angle γ and the cone-angle η specify a particular detector element on the curved detector panel. R_f is the distance between the x-ray source and the isocenter. The gantry rotates around the z axis, which defines the isocenter. R_d is the distance between the isocenter and the center of the detector array. z_{feed} is the axial distance traveled by the patient bed in one rotation.

1.2 CT Data Acquisition

X-ray CT imaging uses an X-ray tube as a source. As the X-rays travel through the patient (or other objects being scanned), the photons are either absorbed by the tissue, scattered (i.e. Compton scattering), or transmitted [13]. As the source and detectors together move around in a certain pattern, the detectors, which are arranged on the opposite side of the patient from the source, collect photons that successfully travel through the patient from different positions and view angles. The proportion of transmission basically depends on the spectrum of the X-ray source, the geometry of the system and the spatially-variant attenuation coefficients of the body. With knowledge of the source and geometry, the measurements are preprocessed and used for the subsequent reconstruction.

1.3 Reconstruction

X-ray CT reconstruction is the procedure that uses the projections collected by the detector array to produce a series of cross-sectional images or a three-dimensional image volume. Reconstruction algorithms for CT reconstruction can be roughly divided into two main categories: analytical algorithms and statistical algorithms. Analytical algorithms are based on a deterministic model for the projections and generally are time-saving since no iterations are required in their implementations, but they may require a larger amount of data to weaken the artifacts. Statistical algorithms, on the other hand, are based on a physical model that reflects the mean and statistical features of the measurements and then optimize an objective function which finds the most likely image given the measured data. A prior model for the image can be included optionally. In this thesis, the analytical helical Feldkamp (FDK) algorithm is introduced first. Next, two statistical algorithms, the alternating minimization (AM) algorithm derived by O'Sullivan and Benac [14] and a different version used for incomplete projections derived by Snyder, O'Sullivan, Murphy, Politte, Whiting and Williamson [15], are illustrated.

1.4 Data Used in this Thesis

Patient data acquired by Dr. Steven Don from a Siemens Sensation 16 scanner located at St. Louis Children's Hospital were to evaluate the sampling strategies investigated in this thesis. A standard abdominal imaging protocol with contrast agent was used. The data were retrospectively processed with full approval from the Washington University Institutional Review Board (IRB). Each dataset contains 1160 views/rotation, approximately 12 rotations, 672 detectors/row and 16 rows. The width of the detectors is 0.00135413 (radius). The distance between the x-ray source and the isocenter, R_f, is 570.0 mm. Rotation time equals to 0.5s. z_{feed} for one rotation equals to 24mm [15].

Chapter 2

Background

This chapter describes the derivation of the FDK algorithm, the AM algorithm and the AM algorithm when projection data are incomplete. Also, the implementations are introduced and some results using real data are shown.

2.1 Feldkamp (FDK) Algorithm

2.1.1 Fourier-Slice Theorem

The Fourier-slice theorem explains the important relationship between the 1-D Fourier transform of a projection and the 2-D Fourier transform of the object. This is the basis of the filtered backprojection method [13].

Suppose we have a projection at angle θ, denoted as $g(l,\theta)$, where l denotes the lateral position of the projection. The 1-D Fourier transform of it can be written as

$$G(\zeta,\theta) = \mathcal{F}_{1D}\left[g(l,\theta)\right] = \int_{-\infty}^{\infty} g(l,\theta)e^{-j2\pi\zeta l}dl, \text{ where } j=\sqrt{-1}.$$

where ζ is spatial frequency. The projection $g(l,\theta)$ can be expressed as

$$g(l,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - l)dxdy.$$

By substituting this into the previous expression, we have

$$G(\zeta,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - l)e^{-j2\pi\zeta l}dxdydl,$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi\zeta(x\cos\theta + y\sin\theta)}dxdy.$$

Now, from the definition of the two-dimensional Fourier transform, we have the important relationship

$$G(\zeta,\theta) = F(\zeta\cos\theta, \zeta\sin\theta).$$

This relationship states that the 1-D Fourier transform of a projection is a slice of the 2-D Fourier transform of the object.

2.1.2 Standard Filtered Backprojection (FBP) Algorithm [13]

The standard FBP algorithm using parallel-beam geometry can be derived using the Fourier-slice theorem. First, the inverse Fourier transform of

$$G(\zeta,\theta) = F(\zeta\cos\theta, \zeta\sin\theta)$$

can be written as $f(x,y) = \int_0^{2\pi}\int_0^{\infty} F(\zeta\cos\theta, \zeta\sin\theta)e^{j2\pi\zeta(x\cos\theta+y\sin\theta)}\zeta\, d\zeta\, d\theta.$

By applying the Fourier-slice theorem, we have

$$\begin{aligned} f(\mathrm{x},\mathrm{y}) &= \int_0^{2\pi}\int_0^{\infty} G(\zeta,\theta)e^{j2\pi\zeta(x\cos\theta+y\sin\theta)}\zeta\, d\zeta\, d\theta, \\ &= \int_0^{\pi}\int_{-\infty}^{\infty} G(\zeta,\theta)e^{j2\pi\zeta(x\cos\theta+y\sin\theta)}|\zeta|\, d\zeta\, d\theta, \\ &= \int_0^{\pi}\left[\int_{-\infty}^{\infty} G(\zeta,\theta)e^{j2\pi\zeta(x\cos\theta+y\sin\theta)}|\zeta|\, d\zeta\right] d\theta. \end{aligned}$$

Here, the projection $g(l,\theta)$ is filtered by a ramp filter whose frequency response is $|\zeta|$. In practice, the high-pass ramp filter should be apodized (cut off in the high frequency region) since the high frequency signal usually contains much noise that should not be amplified.

2.1.3 FDK Algorithm Overview

The FDK algorithm is a three-dimensional generalization of the standard filtered backprojection (FBP) algorithm [12][16][17]. Since a helical geometry and fan-beam are used in 3-D helical CT, applying FBP to a helical geometry requires a rebinning process in two directions:

- rebinning from fan-beam geometry to parallel-beam geometry,

- rebinning from the 3-D data to the approximately equivalent 2-D projections in transverse planes.

Once the projections are in a 2-D parallel geometry in transverse planes, the standard FBP algorithm can be readily applied.

2.1.4 Helical Feldkamp (FDK) Implementation [12]

The basic steps of our implementation are as follows:

1. Perform row-wise fan-beam-to-parallel-beam rebinning.

2. Filter the measured data with a ramp-filter multiplied by either of these two basic apodization windows in the frequency domain:

-Hamming window: $0.54 + 0.46\cos(\pi\omega)$

-Hann window: $0.5 + 0.5\cos(\pi\omega)$

3. Sequentially reconstruct each slice by backprojecting a symmetric set of views on both sides of that slice. The overall backprojection expression is

$$\mu(\mathrm{x},\mathrm{y},\mathrm{z}) = \frac{\pi}{2\theta_m}\int_{-\theta_m}^{\theta_m} \frac{R_f}{\sqrt{R_f^2+\hat{\eta}^2}}\,\omega_{3d}(\theta,\hat{t},\hat{\eta})\tilde{\rho}(\theta,\hat{t},\hat{\eta})d\theta,$$

where $\omega_{3d}(\theta,\hat{t},\hat{\eta})$ is the redundancy weighting function, $\tilde{\rho}(\theta,\hat{t},\hat{\eta})$ is the radially-filtered projection data, and $[\theta_m,-\theta_m]$ is the angular extent of the fan beam.

2.1.5 Results

Some of the reconstructions of the Siemens Sensation 16 data using the Feldkamp algorithm are shown in Figure 2.1. Since the reconstructions are three dimensional, slices from different directions are shown.

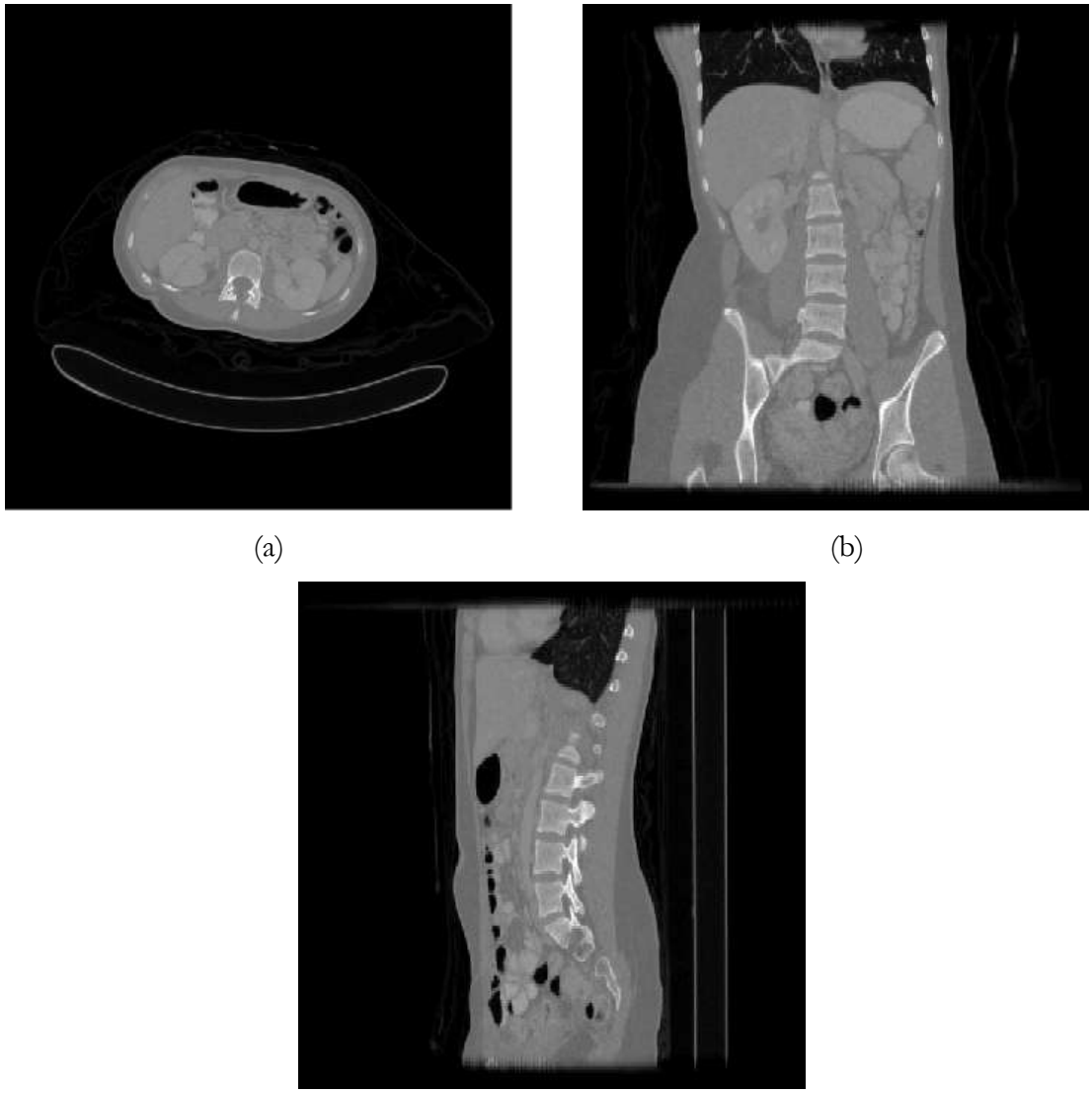


Figure 2.1 Image reconstruction of data from subject number 1 on a $512\times512\times164$ image space using the FDK algorithm. 1160 views were taken. (a) shows the 82nd axial slice , (b) shows the 256th coronal slice, and (c) shows the 256th sagittal slice. The display windows are $[0, 0.0418]$ mm^{-1}.

2.2 Alternating Minimization (AM) Algorithm

2.1.1 AM Algorithm Overview

The AM algorithm for transmission tomography, derived by O'Sullivan and Benac, is an algorithm for finding the maximum-likelihood estimate of the image under the assumption that the measurements obey a Poisson data model [13]. We define the model for the measurements in the following way.

Let $x \in \mathcal{X}$ be a coordinate in the image space and $y \in \mathcal{Y}$ be a source location and detector pair in the measurement space. The measured data $d(y)$ are mutually independent and Poisson distributed given $c(x)$, where $c(x)$ is the attenuation coefficient to be estimated. Therefore

$$\Pr[d(y) = n] = \frac{q^n(y:c)}{n!} e^{-q(y:c)},$$

in which

$$q(y:c) = I_0(y) \exp\left[-\sum_{x \in \mathcal{X}} h(y \mid x) c(x) \right], \quad \text{(Beer's Law of Attenuation)}$$

where $I_o(y)$ is the source intensity at the source-to-detector pair y and $h(y \mid x)$ is the scanner's point-spread function. The summation in the exponent represents the forward projection of the attenuation coefficients.

The maximum-likelihood estimate of the image is found by solving the minimization problem

$$\hat{c}(x) = \arg\min_{c \in \mathcal{C}} I[d(y), y \in \mathcal{Y} \mid q(y:c), y \in \mathcal{Y}], \qquad x \in \mathcal{X}$$

where

$$I[d(y), y \in \mathcal{Y} \mid q(y:c), y \in \mathcal{Y}] = \sum_{y \in \mathcal{Y}} d(y) \ln\left[\frac{d(y)}{q(y:c)} \right] - d(y) + q(y:c).$$

The alternating-minimization (AM) algorithm identified by O'Sullivan and Benac (2006) produces a sequence of estimates with non-increasing I-divergence (non-decreasing likelihoods) [13]:

$$\hat{c}_{\mathrm{AM}}^{(k+1)}(x) = \max\left[\hat{c}^{(k)}(x) - \frac{1}{Z(x)}\ln\left[\frac{\sum_{y\in\mathcal{Y}} h(y\,|\,x)d(y)}{\sum_{y\in\mathcal{Y}} h(y\,|\,x)q(y:\hat{c}^{(\mathrm{k})})}\right], 0\right],$$

where $Z = \max_{y}\sum_{x\in\mathcal{X}} h(\mathrm{y}\,|\,\mathrm{x})$ and k denotes the iteration index.

2.1.2 AM Algorithm Implementation

The basic steps of AM implementation using HECTARE 1.0 for our experiments are shown as follows.

1. Allocate space for the current image estimate, the transmission data, two backprojections and one forward projection, read in parameters, transmission data, *Bowtie* file and I_o file, and save them in the formats used by HECTARE 1.0.

2. Calculate the update step size $1/Z$ based on $Z = \max_{y}\sum_{x\in\mathcal{X}} h(\mathrm{y}\,|\,\mathrm{x})$. Restricting the step size in this manner ensures convergence [13].

3. If no data subsets are used, in each iteration forward-project the current image estimate to the detector space, save as *Meandata*.

4. Perform exponentiation based on: $Meandata = \exp(Meandata)\times Bowtie\times I_o$.

5. Back-project *Meandata* and *Transmissiondata* to image space, save as *Meanbackprojection* and *Databackprojection* respectively.

6. Update current image estimate based on:

$$\mathrm{current} = \max\left[\mathrm{previous} - \frac{1}{Z}\ln\left[\frac{Databackprojection}{Meanbackprojection}\right], 0\right].$$

7, Write out the final estimate after all iterations are run.

2.1.3 Results

Data from a second subject scanned on the Siemens Sensation 16 were reconstructed using the AM algorithm. Some representative image slices are shown in Figure 2.2. (All subsequent reconstructions and figures in this thesis will use the data from this second subject.)

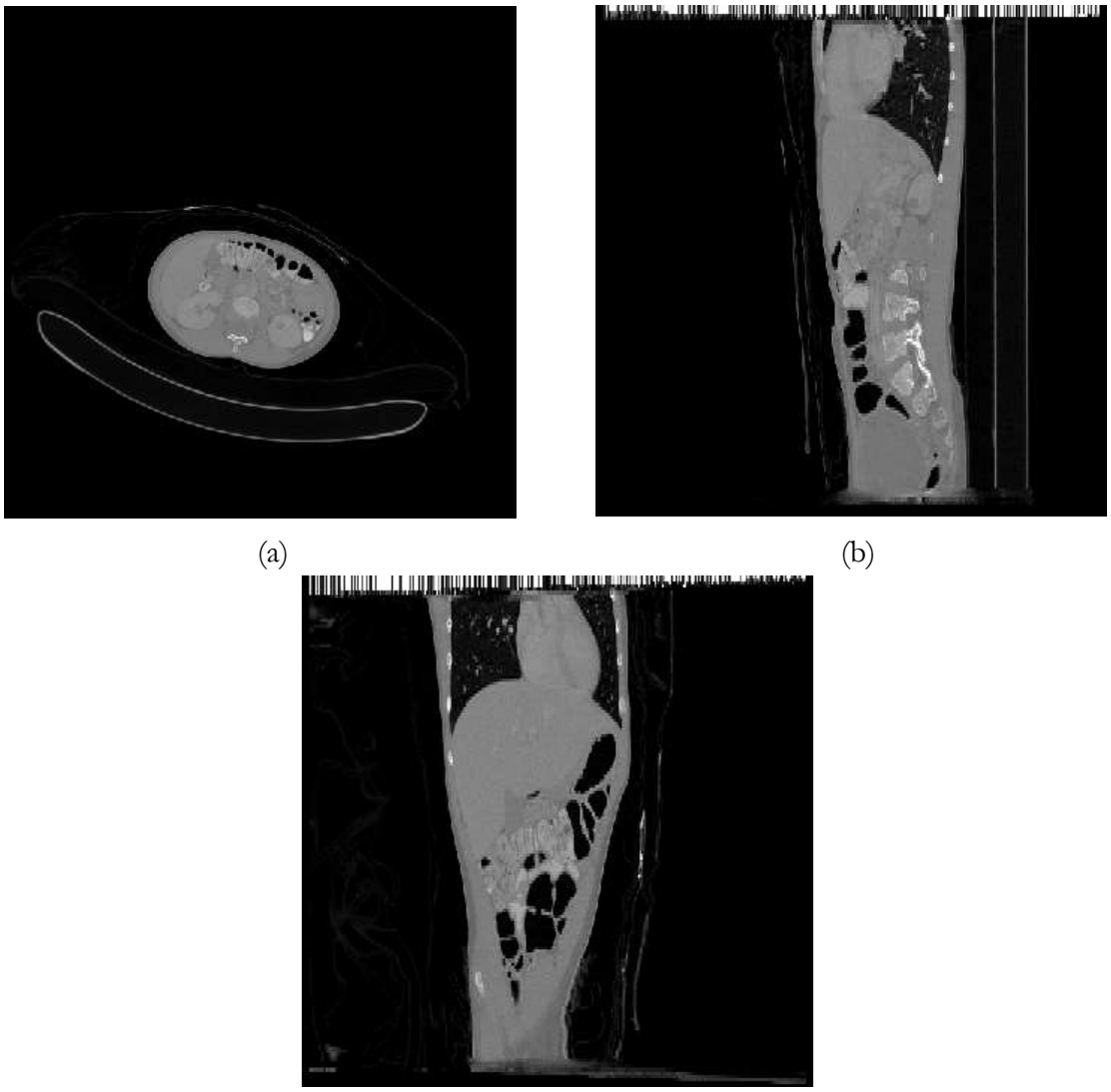


Figure 2.2 Image reconstruction of data from subject number 2 on a $512\times512\times164$ image space using the AM algorithm. 1160 views were taken. (a) shows the 82nd axial slice , (b) shows the 256th coronal slice, and (c) shows the 256th sagittal slice. The display windows are $[0, 0.0418]$ mm^{-1}.

2.3 Alternating Minimization (AM) Algorithm for Incomplete Projection Data

2.3.1 AM Algorithm for Incomplete Projection Data Overview

The reconstruction problem is to solve the minimization [14]

$$\hat{c}(x) = \arg\min_{c \in \mathcal{C}} I\left[d(y), y \in \mathcal{Y}_{\text{inc}} \mid q(y:c), y \in \mathcal{Y}_{\text{inc}}\right],$$

which is analogous to the previous problem, but we utilize only the data from the set $y \in \mathcal{Y}_{\text{inc}}$. There is no twist here; the algorithm uses all the available projections and finds the most likely image by minimizing the I-divergence between the measurements and the forward projection of the image being estimated. The minimization is equivalent to maximizing the log-likelihood [1].

The sequence of estimates can be defined by

$$\hat{c}_{\text{AM}}^{(k+1)}(x) = \max\left[\hat{c}^{(k)}(x) - \frac{1}{Z(x)} \ln\left[\frac{\sum_{y \in \mathcal{Y}_{\text{inc}}} h(y \mid x) d(y)}{\sum_{y \in \mathcal{Y}_{\text{inc}}} h(y \mid x)\, \mathrm{q}(y:\hat{c}^{(k)})}\right], 0\right].$$

2.3.2 AM Algorithm for Incomplete Projection Data Implementation

The implementations of the AM algorithm for incomplete projection data are similar to the original AM algorithm explained in 2.1.2. For completeness, the basic steps of the AM algorithm for incomplete projection data are explained as follows.

1. Allocate space for an indicator, the current image estimate, the transmission data, two backprojections and one forward projection, read in parameters, transmission data, bowtie file and I_o file, and save them in the formats used by HECTARE 1.0. The indicator is a matrix of zeros and ones that specifies which detectors are in the set $\mathcal{Y}_{\text{inc}}$.

2. Calculate the update step size $1/Z$ based on $Z = \max_y \sum_{x \in \mathcal{X}} h(\mathrm{y} \mid \mathrm{x})$ to ensure convergence [1].

3. If no data subsets are used, in each iteration, forward-project the current image estimate to the detector space, save as $Meandata$.

4. Perform exponentiation based on: $Meandata = \exp(Meandata) \times Bowtie \times I_o \times Indicator.$

5. Set the projection to 0 at the positions where the detectors are not used in the simulation: $Transmissiondata = Transmissiondata \times Indicator.$

6. Back-project $Meandata$ and transmission data to image space, save as $Meanbackprojection$ and $Databackprojection$ respectively.

7. Update current image estimate based on:

$$\text{current} = \max\left[\text{previous} - \frac{1}{Z}\ln\left[\frac{Databackprojection}{Meanbackprojection}\right], 0\right].$$

8. Write out the final estimate out after all iterations are run.

Some results for this algorithm using different sampling strategies with the real data will be shown in Chapter 3.

2.4 Adding an *a Priori* Model for the AM Algorithm

2.4.1 Overview

To suppress noise in the image arising from noise in the data, we penalize the difference between neighboring voxels, which enforces smoothness on the reconstructed image [15][18][19]. One such regularization penalty can be

$$R(c) \triangleq \sum_{x \in \mathcal{X}} \sum_{x' \in N_x} \omega_{x,x'} \psi\left(c(x) - c(x')\right),$$

where N_x is the set of the 26-voxel neighborhood surrounding voxel x and $\psi(\cdot)$ is the potential function.

The neighborhood weights, $\omega_{x,x'}$, are chosen to be the inverse distance between voxel centers. $\psi(\cdot)$ is a symmetric convex function which can be written as

$$\psi(\mathrm{t}) \triangleq \frac{1}{\delta^2}\left[|\delta t| - \ln(1+|\delta t|)\right],$$

where δ is a parameter that controls the transition between a quadratic region (for smaller t) and a linear region (for larger t).

The complete cost function $\Phi_{\mathrm{AM}}(c)$, including the regularization penalty, $R(c)$, is

$$\Phi_{\mathrm{AM}}(c) = I[d(y), y \in \mathcal{Y} \mid q(y:c), y \in \mathcal{Y}] + \beta R(c).$$

Our goal is to minimize $\Phi_{\mathrm{AM}}(c)$ by iteratively updating the image estimate $c(x)$. D. B. Keesing and J. A. O'Sullivan dan updating strategy which decouples the objective function and iteratively approach the local minimum by implementing Newton's Method. The detailed deviation of the regularized AM algorithm can be found in [12] and [15]. The derivation is also attached in Appendix A for convenience. Some of the notation has been changed from that in the references for consistency of this thesis.

When the objective function includes the penalty term, the estimated image is said to be "regularized." We will use the terms "penalized" and "regularized" interchangeably below.

2.4.2 Newton's Method with Trust Region Modification

To minimize the decoupled penalized objective function, Newton's method is a good choice since the optimization here is a one-dimensional convex optimization problem with no constraints. Generally, for most points that are estimated in our case, only a few iterations of Newton's method are needed to get close enough to the local minimizer.

The well-known Newton's method for solving the unconstrained minimization problem can be written as

$$x^{k+1} = x^k - G^{-1}(\mathrm{x}^k) f'(\mathrm{x}^k),$$

where $f'(\mathrm{x}^k)$ is the gradient at the current estimate and $G(\mathrm{x}^k)$ is the Hessian matrix. The most fundamental requirement for Newton's method is that the function f has to be twice differentiable everywhere in the function's domain. However, in general, successful implementation of Newton's

method for practical applications requires further conditions. First, Newton's method has no guarantee of convergence, especially when the initial guess is not "close" enough to the local minimizer. (In our case, the local minimizer is also a global one since our decoupled objective function is convex for $x \in R^n$.) Second, the Hessian matrix may be singular (not invertible) so that no explicit solution can be given. Third, the Hessian matrix may be indefinite (e.g. at the saddle point), which makes some updating steps nonsensical [20].

In our minimization problem, nonsensical steps might be taken by Newton's method when the image update at a certain step depends on the evaluation of the Hessian in an almost linear region of the objective function. Under these conditions, the quadratic surrogate is not a good approximation of the objective function, resulting in an update that is too large. The image update might either oscillate back and forth (probably coming down to the local minimizer eventually) or generate NaN or Inf when the value exceeds the maximum. (There is a maximum since the implementations are done in C++.) Experience with numerous numerical experiments has shown that the probability of a NaN being computed rises when less data is used in the reconstruction. This is so because under these conditions the edge-preserving penalty would go to the linear region more readily.

The idea of adding a trust region to force the convergence of Newton's method is intuitive; we only "trust" updates within the trust region and otherwise skip that update. At the same time, we either keep, shrink or expand the trust region based on the decreasing rate of change of our objective function. The version of Newton's method we used is based on [20].

The algorithm for our optimization can be illustrated in this way:

(1) Initialize parameters: set $k = 1$, $k_{\max}$ be the maximum number of iterations, $0 < \gamma_1 < 1 < \gamma_2$ and γ_3 be a small number between 0 and 1 (0.05 in my experiments). Set $0 < \eta_1 < \eta_2 < 1$. Set radius rad and the maximum of radius rad_max to any positive values.

(2) Calculate the value of our decoupled objective function ($f(\mu)$) and its first ($f'(\mu)$) and second order derivatives ($f''(\mu)$). This calculation is the same as for Newton's method illustrated previously. Formulas can be found in Appendix A.

(3) If converge, then STOP. The convergence criterion is set up as:

(A) when $decrement$ drops below a small enough value,

where $decrement = f'(\mu)^2 / \left[2 \times f''(\mu)\right]$.

(B) when rad drops below a small enough value.

(C) when $k = k_{\max}$.

When any of the above criteria are satisfied, the algorithm stops.

(4) Calculate Newton's method step using

$$w = -f'(\mu) / f''(\mu).$$

If the absolute value of w exceeds the boundary trust region, set w to the boundary.

(5) Calculate the new value of our decoupled objective function and its first and second order derivatives.

(6) Calculate the actual reduction of the objective function value and the predicted reduction.

$$\mathrm{RE}_{actual} = f(\mu) - f(\mu + \omega),$$

$$RE_{predict} = -f'(\mu) \times w - w^2 \times f''(\mu) / 2.$$

$$ratio = \frac{\mathrm{RE}_{actual}}{\mathrm{RE}_{predict}}.$$

(7) If $ratio < \eta_1$, then $rad = \gamma_1 \times rad.$

If $ratio > \eta_2$, then $rad = \min(\gamma_2 \times rad, rad\_max).$

If $f(\mu + \omega) = \infty$, then $rad = \gamma_3 \times rad.$

(8) If $\mathrm{RE}_{actual} > 0$, update image and iteration:

$$\mu = \mu + \omega,$$

$$k = k + 1.$$

(9) Go to step 3.

End.

Comments:

1. There are ways to update the trust region radius. For example, γ_1, γ_2, γ_3 can be adaptively decided based upon available information. This idea is mentioned in [20].
2. To predict the reduction of our objective function, a quadratic surrogate is introduced at the current μ. The quadratic surrogate at μ is defined by the first three terms of the Taylor series for $f(\mu)$. The original Newton's method takes the minimizer of the quadratic surrogate as its

updating step. With trust region modification, we take the regular Newton's step only if that step is within the trust region. Otherwise, we use the boundary of our trust region.

3. There are some possible modifications that can be done here, which might make the algorithm converge with lower computational complexity in some cases. For example, in step 7, if $ratio < \eta_1$, we shrink ω directly by a factor of γ_1 and check only the ratio. We keep shrinking it until the ratio goes to a value larger than η_1. Since we only need to recalculate the function value at $(\mu + \omega)$ without computing the derivatives, this would probably reduce some computation in some cases.
4. The convergence of this algorithm has been shown by many authors including D. C. Sorensen [20].

2.4.3 Implementation

In the original HECTARE, a regular Newton's method is implemented. However, when we "throw away" a fraction of data, Newton's method has no guarantee of convergence and may generate NaNs (not a number) when implemented. This is the original motivation for implementing a modified Newton's method for our case to force convergence and avoid the NaN issue.

Images computed using unregularized AM and regularized AM with different penalty weights are shown in Figure 2.3. Based on the profiles of values of the image in the highlighted row of pixels, the larger penalty weight is seen to correspond to a smoother image, but without any apparent blurring of the edges of transition regions.

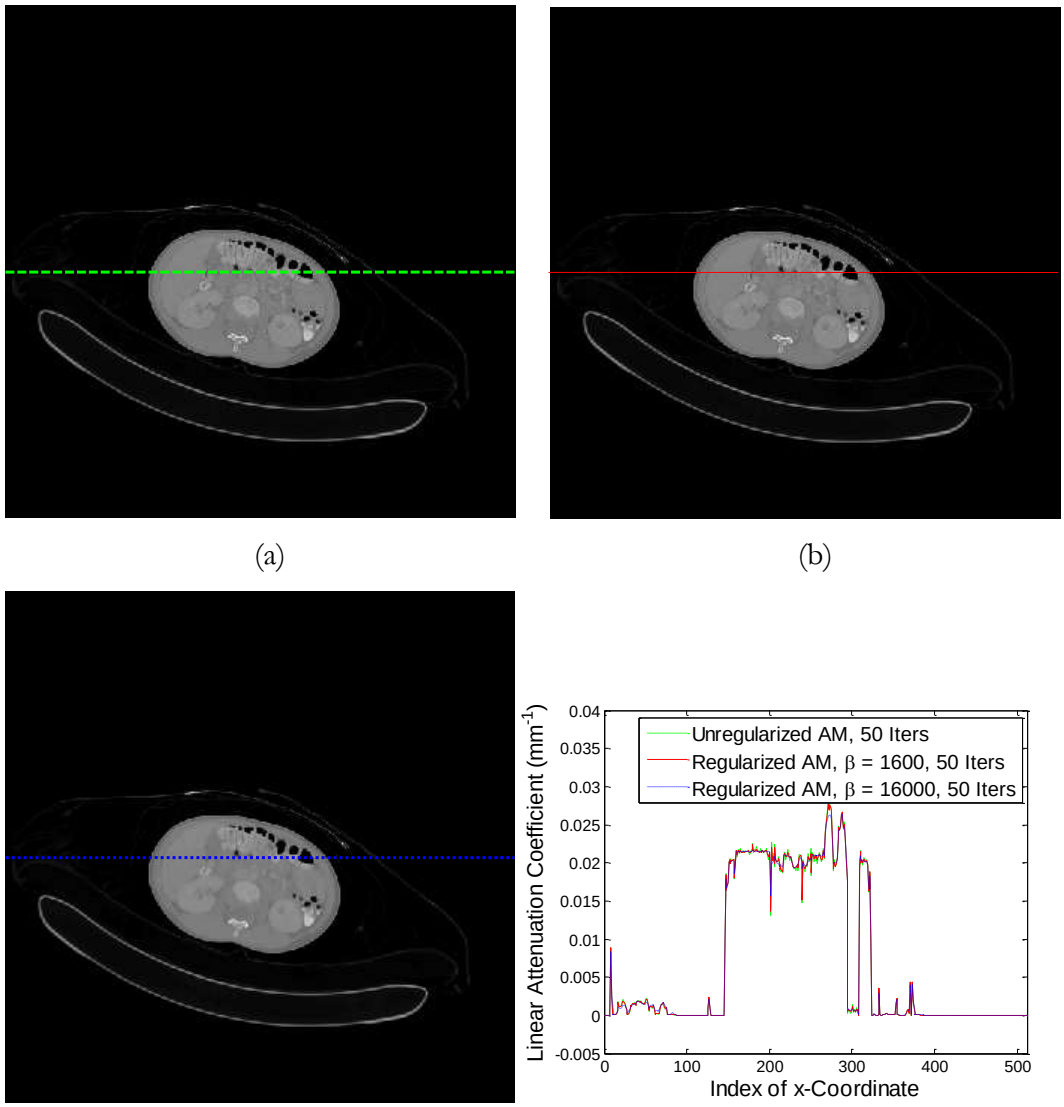


Figure 2.3 Image reconstruction on a $512\times512\times164$ image space using the AM algorithm. The 82nd axial slice is shown in all panels. (a) unregularized reconstruction, 50 iterations, (b) regularized reconstruction, penalty weight $\beta=1600$, 50 iterations, (c) regularized reconstruction, penalty weight $\beta=16000$, 50 iterations, 82nd and (d) the image profiles through row 256. The display windows are $[0, 0.0418]$ mm-1.

2.5 Ordered Subsets

Ordered subsets (OS), also known as a block iterative or incremental gradient method, is a well-known technique used to speed up the convergence of a large family of iterative algorithms [21]. Due to the high computational intensity, iterative algorithms are not widely used in clinical applications. However, techniques like ordered subsets can significantly increase the convergence speed. OS are now widely used. One of the famous examples is the EM algorithm for emission tomography, which usually introduces OS.

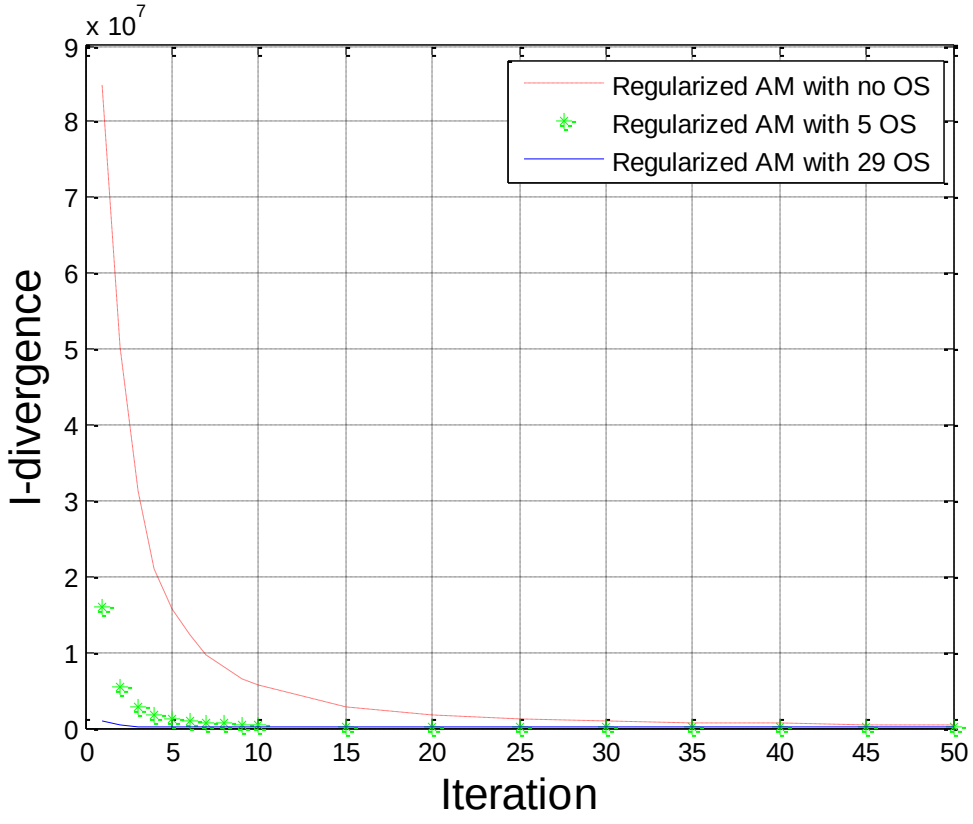
The main idea of the OS method is that, for each iteration, one can divide the measurements into ordered disjoint sets and run the algorithm with data in each subset sequentially. The advantage is obvious: in each iteration, when executing the algorithm with the n^{th} subset, the object has already been updated with the previous (n-1) subsets. In the CT application, the computational cost of executing forward and backprojection does not increase when OS are used since the number of source-detector pairs to be projected stays the same as in the regular AM iteration, no matter if these projections are performed for every measurement in the context of an iteration of an ordinary algorithm or if subsets of the measurements are projected within the "subiterations" of the OS version of the algorithm. Although the computational cost of executing the image update increases proportionally with the number of ordered subsets, this cost is lower than the cost of projection operations in general. This is especially the case when executing the regular AM algorithm, in which the updates are analytical and the computational costs for this operation are very low.

However, it has to be pointed out that ordinary OS algorithms generally do not converge to the optimal point. Reference [26] offers an OS algorithm in which convergence is guaranteed by the introduction of relaxation parameters. This paper also mentions that convergence is more important in regularized methods because, for maximum likelihood (ML) methods, one usually does not run ML until convergence. Moreover, for most cases, there is usually a trade-off between speed acceleration and image quality with respect to the number of subsets.

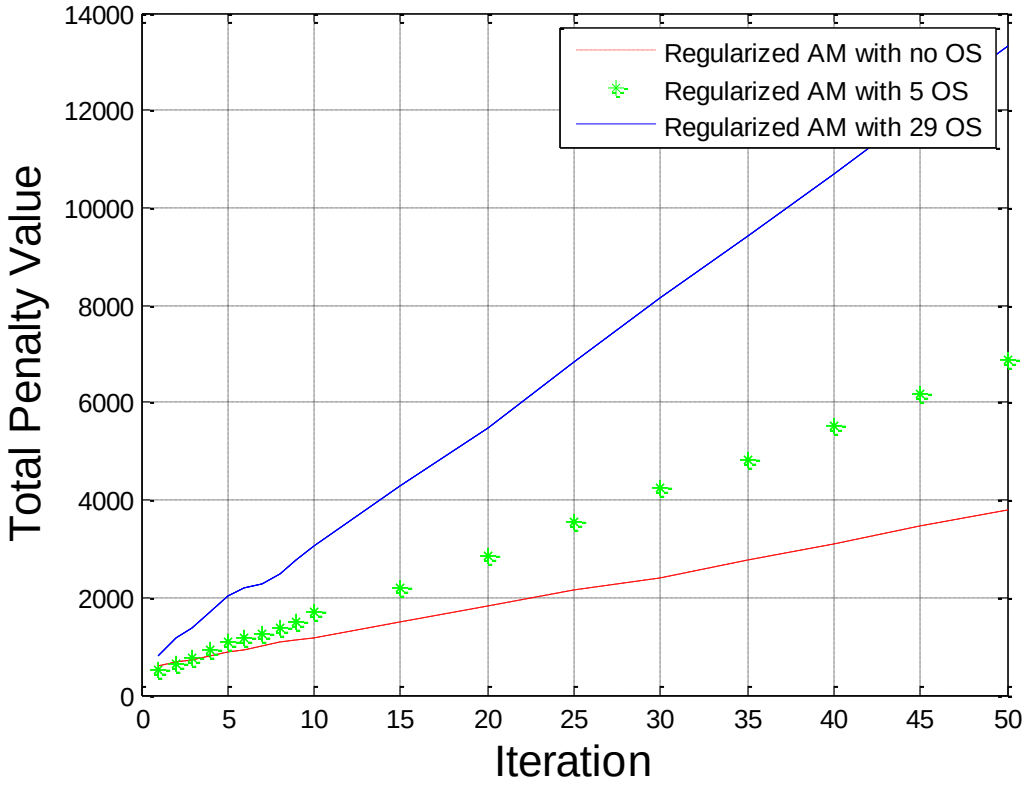
In addition, some other papers [21]-[24] also proposed the use of an aggressive number of subsets in the first few iterations and then to switch to fewer subsets after that. However, in this thesis, we

simply choose a reasonable number of subsets and never switch it since the main concern here is about sampling strategies. I did a lot of experiments to adjust the number of subsets in different cases by comparing the cost function values to achieve a balance. The images presented in the following sessions are at least "sub-optimal".

To investigate the effect of including OS on the convergence behavior of the AM algorithm, regularized AM algorithm with different numbers of subsets are tested and objective function value at different iterations are tracked. In Figure 3.1, we compare the values of the I-divergence term and penalty term separately.



(a)



(b)

Figure 2.4 Image reconstruction on a $512\times512\times164$ image space with full data using the regularized AM algorithm. (a) a plot of I-divergence at different iterations with different number of subsets, (b) a plot of total penalty values at the corresponding iterations with the corresponding number of subsets.

The result shows that with more subsets, the data fitting term, I-divergence, in regularized AM algorithm almost converges to the same point with 1, 5, 29 OS being used. The convergence is speeded up almost proportional to the number of OS. As for the penalty term, it diverges faster with more OS.

Chapter 3

Sampling

Sampling is the process of converting a signal, which in general is continuous, into a numeric sequence (discrete time or space). In this chapter, we mainly focus on sampling issues in the CT application.

3.1 Nyquist-Shannon Sampling Theory Applied to CT

The famous Nyquist-Shannon sampling theorem states that if a function $x(t)$ contains no frequencies higher than B Hertz, it is completely determined by giving its ordinates at a series of points spaced $1/(2B)$ seconds apart. In other words, a bandlimited function can be perfectly reconstructed from a countable sequence of samples if the bandlimit, B, is no greater than half the sampling rate. The following pages contain analysis of the Nyquist-Shannon sampling theory in the CT application.

3.1.1 Number of Samples per Projection

The first goal is to estimate the number of samples per projection, N_s needed in CT, by applying the Nyquist-Shannon theorem. Setting $f_{\max}$ as the highest frequency or bandwidth limit of the imaging system, the theorem specifies that the sampling frequency $f_s \geq 2 f_{\max}$ [25]. Now the number of samples per projection based on the sampling theorem is

$$N_s = 2 f_{\max} \times FOV ,$$

where FOV is the diameter of the field of view to be imaged.

Specifically, for most of the experiments shown in this thesis, each image slice is set to be 512×512 pixels and each pixel is $1mm \times 1mm$. In our case, the Siemens Sensation 16 scanner has $N_s = 672$. Therefore, if the maximum frequency is lower than $0.656/mm$, the reconstruction can avoid aliasing. Since the pixel size in the transverse (axial) plane is $1mm \times 1mm$, the maximum resolution due to this choice is $0.5/mm$.

3.1.2 Number of Projections

The number of projections, N_p, is calculated to ensure complete filling of the k-space up to the highest frequency [25]. The Fourier-Slice theorem, illustrated previously in chapter 2, states that the 1-D Fourier transform of a projection is a slice of the 2-D Fourier transform of the object. In other words, the 1-D Fourier transform of a projection equals a line passing through the origin of the 2-D Fourier transform of the object at the angle corresponding to that projection [13]. Therefore, for $N_s = 2 f_{\max} \times FOV$, to accomplish the complete filling, the number of samples/distance along the largest circle in u, v space (see Figure 3.1) has to be set equal to or larger than the number of samples/distance along the radial direction [25]. Here, N_s determines the dimension of the corresponding Fourier space.

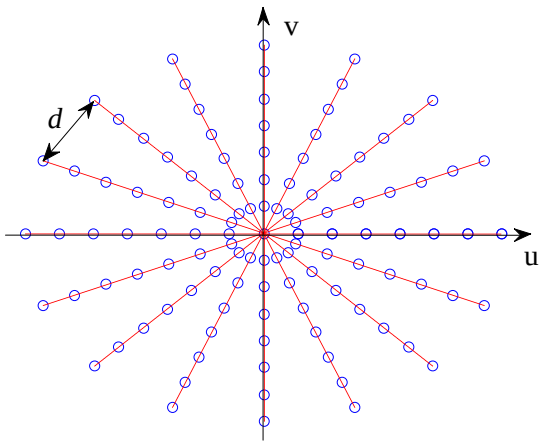


Figure 3.1 The K-space of the image domain is shown. The points on each line are equivalent to the Fourier transform of the corresponding projection. *d* denotes the distance between any adjacent points on the largest circle. Based on our analysis, d has to be smaller than the distance between any adjacent points on any line.

Now, since the number of points in the diameter of the largest circle in k-space is N_s and each projection provides two points along the circumference of the largest circle, the minimum number of projections needed is $N_p = (\pi/2) \cdot N_s$. Specifically, the CT system we used for our experiments contains 672 detectors per row and therefore at least 1056 views are required. The number of views of the Siemens Sensation 16 is 1160, which is enough based on the analysis above.

3.2 Compressed Sensing

3.2.1 Introduction to Compressed Sensing

The Fourier-slice theorem indicates some relationships between the projections in the data space and the object in the image space and helps us to understand how angular sampling can influence reconstructions[19]. On the other hand, conventional approaches to sampling signals follow the celebrated Nyquist-Shannon sampling theory: the sampling frequency must be at least twice the maximum frequency of the signal so that it can be perfectly recovered [5][6][10]. In practice, this is the fundamental principle for numerous applications, ranging from audio and visual signal sampling (analog-to-digital) to imaging systems. Notice that in many applications, the signal is not band-limited or the highest frequency is so high that too many samples are required. For example, for a period of music played by real instruments, since it is time limited, it is not band-limited. Therefore, the Nyquist-Shannon sampling theory, in this case, cannot be applied directly. To deal with it, the sampling rate is determined by the desired temporal resolution [5] and the higher frequency part of the signal will be treated as noise and filtered out. Specifically, in our medical CT application, the object is not band-limited since it is space limited. In this case, the desired highest frequency for the signal taken into consideration is determined by the number of pixels per millimeter in the field of view. Also, when FBP-based (filtered backprojection) algorithms are implemented, it is common to use an anti-aliasing low-pass filter to bandlimit the signal before sampling.

So far, all the analysis about sampling above in this chapter is based on the Nyquist –Shannon sampling theory. A more novel sensing/sampling paradigm, compressed/compressive sensing/sampling (CS), asserts that one can recover images/signals from far fewer samples or

measurements [5][6][10]. CS relies on the sparsity of the signals. Here, sparsity refers to the idea that many natural signals, either continuous time or discrete time, can be approximated with small error by much less data than is required by the Nyquist-Shannon sampling theory. In other words, many natural signals have sparse representations (most coefficients are small or zeros) when the basis is chosen in a smart way. One of the most common examples is the wavelet transform. Also, if matrix decomposition methods are implemented, one can also force the sparsity by adding a sparsity constraint. These ideas are introduced by numerous papers.

Mathematically, a signal can be represented as

$$f(t) = \sum_{i=1}^{n} x_i \psi_i(t),$$

where $\psi_i(t)$ are the orthonormal basis and x_i are the coefficients.

The signal $f(t)$ is called K-sparse if only K of the x_i are nonzero and $(n-K)$ are zeros. Compressed sensing concerns mostly the case when $K \ll n$. In this case, the signal is highly compressible. Once signals are K-sparse and compressible, one can do transform coding. First, the full n-sample signal $f(t)$ is acquired. After a signal basis $\psi_i(t)$ is selected, one can compute the coefficients x_i via $\boldsymbol{x} = \Psi^T \boldsymbol{f}$. Then, the K highest x_i are encoded and the others are discarded. Now the question is: since $(n-K)$ coefficients are thrown away, why do we sample the full n-sample signal $f(t)$? Can we directly sample the K useful coefficients without acquiring n samples?

The answer is positive. Consider a linear measurement process that has an $m \times n$ sensing matrix $\Phi_{m \times n}$. Suppose the sensing system is underdetermined, that is $m < n$. Now the measurements can be expressed as

$$y = \Phi \boldsymbol{f} = \Phi \Psi \boldsymbol{x}.$$

Now if we can design a stable (non-adaptive) measurement matrix $\Phi_{m \times n}$ such that the down-sampling process ($m < n$) can save the information of the K highest coefficients and then derive a reconstruction algorithm to recover the signal, we would be able to take only m samples but reconstruct higher dimensional signals.

Many papers, including [5], [6], and [10] illustrated three conditions for this measurement matrix $\Phi_{m\times n}$ to be "decent."

First condition (necessary and sufficient condition):

Suppose $m > K$, a necessary and sufficient condition for $\Phi_{m\times n}$ to be "decent" is that, for any vector ρ sharing the same K nonzero entries as $\boldsymbol{x}$ and for some $\varepsilon > 0$,

$$1-\varepsilon \le \frac{\|\Phi\Psi\rho\|_2}{\|\rho\|_2} \le 1+\varepsilon .$$

This means that the matrix $\Phi\Psi$ must be norm-preserving as to the vector ρ.

Second condition (sufficient condition):

A sufficient condition for K-sparse signals is that $\Phi\Psi$ satisfies the first condition for an arbitrary 3K-sparse vector ρ. This condition is called the restricted isometry property (RIP).

Third condition, named the incoherence condition, is related to the second one. Coherence refers to a relationship between the sensing/sampling basis $\varphi_j(\mathrm{t})$ and the signal basis $\psi_i(\mathrm{t})$. The level of coherence can be defined as [5]

$$C[\Phi_i(\mathrm{t}),\Psi_j(\mathrm{t})] = \sqrt{N}\cdot \max_{1\le i,j\le N}\left|\left\{\varphi_j,\psi_i\right\}\right|.$$

Here, $C[\Phi_i(\mathrm{t}),\Psi_j(\mathrm{t})] \in [1,\sqrt{N}]$.

The incoherence condition forces that the rows $\left\{\varphi_j\right\}$ of Φ cannot sparsely represent the columns $\left\{\psi_i\right\}$ of Ψ (and vice versa).

Interestingly, both the RIP condition and incoherence condition can be achieved with high probability by choosing Φ as a random matrix. A simple proof of these conditions can be found in [26]. This conclusion is our motivation for the random sampling experiments.

3.2.2 Sampling Issues in Helical CT

There are numerous kinds of imaging systems being developed and widely used. Examples are traditional film cameras and modern digital cameras. As for computed tomography, the system generates sinograms as the measurements. Each point in a sinogram is an inner product of the image with a source-to-detector line. Those lines define the row vectors in the system matrix and therefore define the "sampling basis." Mathematically, if we define the image as $c(x)$ with size N, the measurements as $d(\mathrm{y})$ with size M and the system matrix A with size $M \times N$. One possible model for the CT system can be

$$\boldsymbol{d} = A\boldsymbol{c} + \boldsymbol{n},$$

where $\boldsymbol{n}$ is additive noise.

In our helical CT system, for a certain detector at a certain position, the data collected by that detector only depends on those voxels cut by that source-to-detector line. So, in our case, the system matrix A is sparse. Motivated by the theorems from compressed sensing (CS), we wish the system matrix A to be a random one. However, with our fixed CT system, no solutions have been found to make the system random enough to satisfy the RIP and incoherence condition. Nevertheless, in uniform-pattern and random-pattern sampling introduced in section 3.3, the experiments were done by adding a random mask on the sinogram (setting some of the values to zero). In the next section, we focus on the experiments and analysis.

3.3 Possible Sampling Strategies in Helical Geometry

3.3.1 An Overview of Our Experiments

Fortunately, with full data available, various sampling approaches can be simulated by using partial data to do the reconstruction. Based on the previous analysis, we tested some subsampling strategies.

1. Subsampling by views:

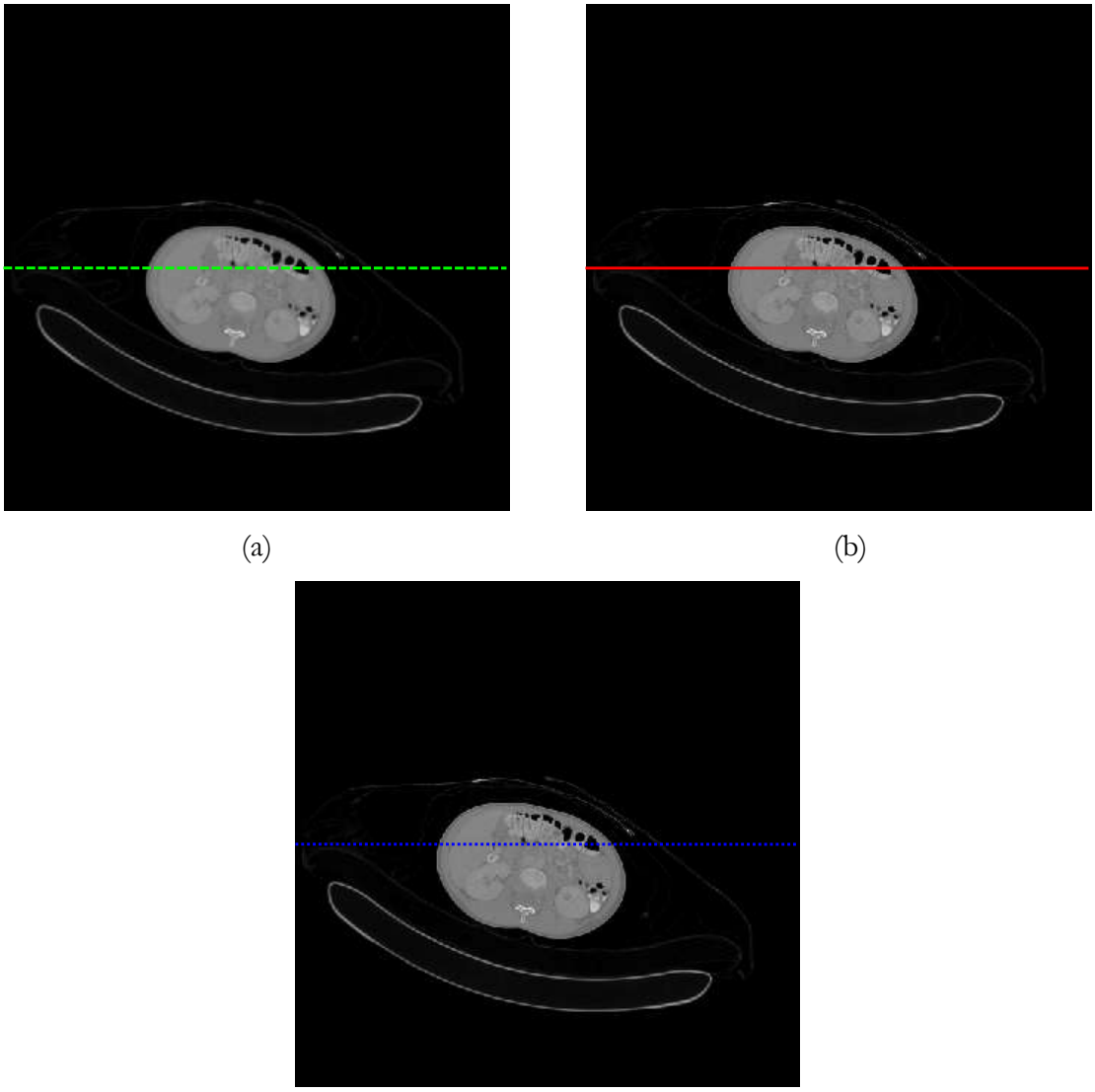
- Uniformly take a subset of the views, i.e., every 2^{nd}, every 5^{th}, every 10^{th} or every 29^{th} view.
- Algorithms used: FDK, unregularized AM and regularized AM.

2. Subsampling by detectors:

- Uniform-pattern sampling: randomly select a subset of detectors for the first view and use the same sampling pattern for all the other views.
- Random-pattern sampling: randomly select a subset of detectors for each view independently.
- Algorithms used: unregularized AM and regularized AM.

3.3.2 Subsampling by Views

In this section, we choose a subset of uniformly spaced views to use in the reconstructions and discard the remaining views. Starting with the full-data case, we progressively discard more and more views and plot the profile through the central cutting line to evaluate the performance.



(a) (b)

(c)

Figure 3.2 Image reconstruction from full-view data, all detectors. The 82nd axial slice is shown in all panels. (a) FDK algorithm, (b) unregularized AM algorithm, 50 iterations, and (c) regularized AM algorithm, penalty weight $\beta = 100$, 50 iterations. The display windows are $[0, 0.0418]$ mm^{-1}.

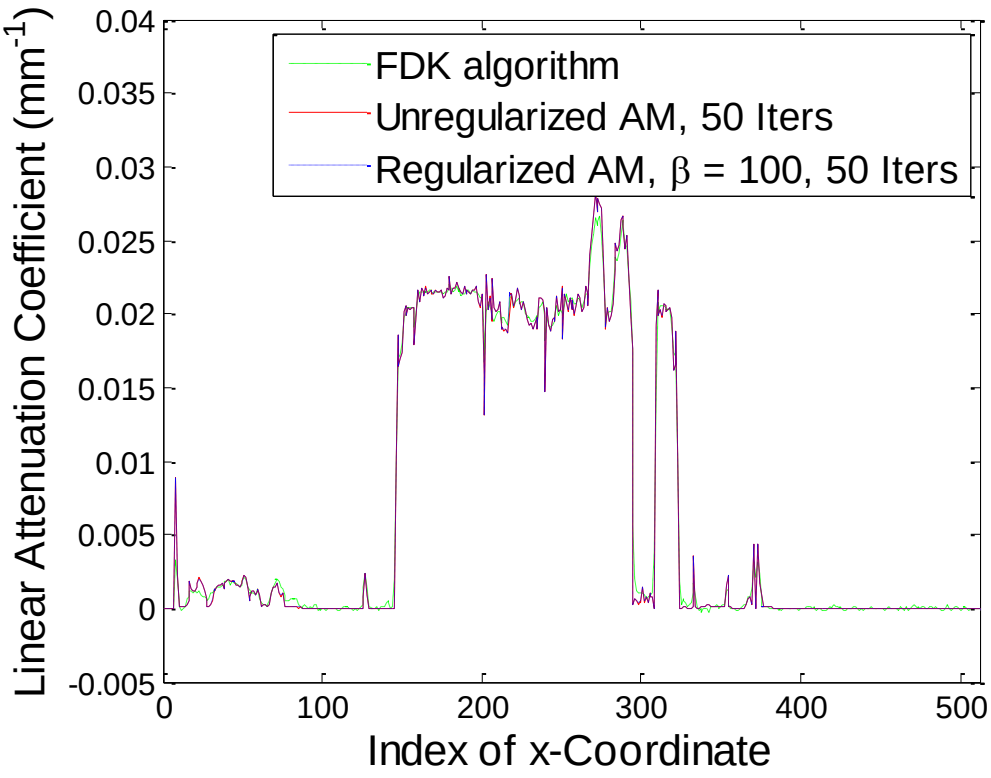


Figure 3.3 Profiles of the FDK, unregularized AM and regularized AM with $\beta = 100$ images shown in Figure 3.2. Full-views are used here. The profiles are plotted along the lines at row 256, as shown in Figure3.2.

In this experiment, with all the data, all of the three algorithms perform well. Figure 3.3 shows that the profiles are close to each other. In the region of air, FDK generates some small fluctuations that are not observable from Figure 3.2(a). In the next experiment, we take every other view and "throw away" the others.

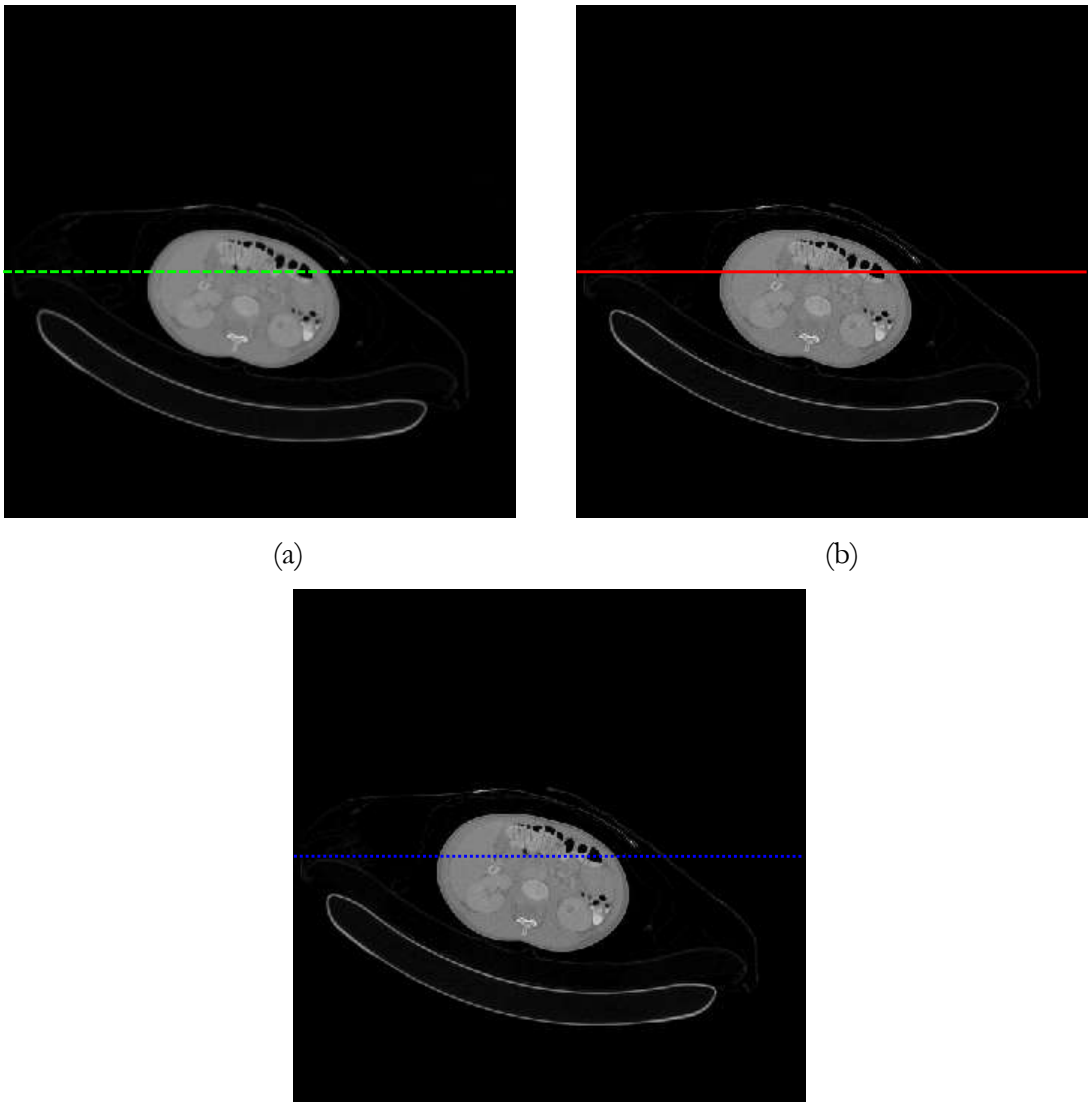


Figure 3.4 Image reconstruction from 1/2 of the views of data, all detectors. The 82nd axial slice is shown in all panels. (a) FDK algorithm, (b) unregularized AM algorithm, 50 iterations, and (c) regularized AM algorithm, penalty weight $\beta = 100$, 50 iterations. The display windows are $[0, 0.0418]$ mm^{-1}.

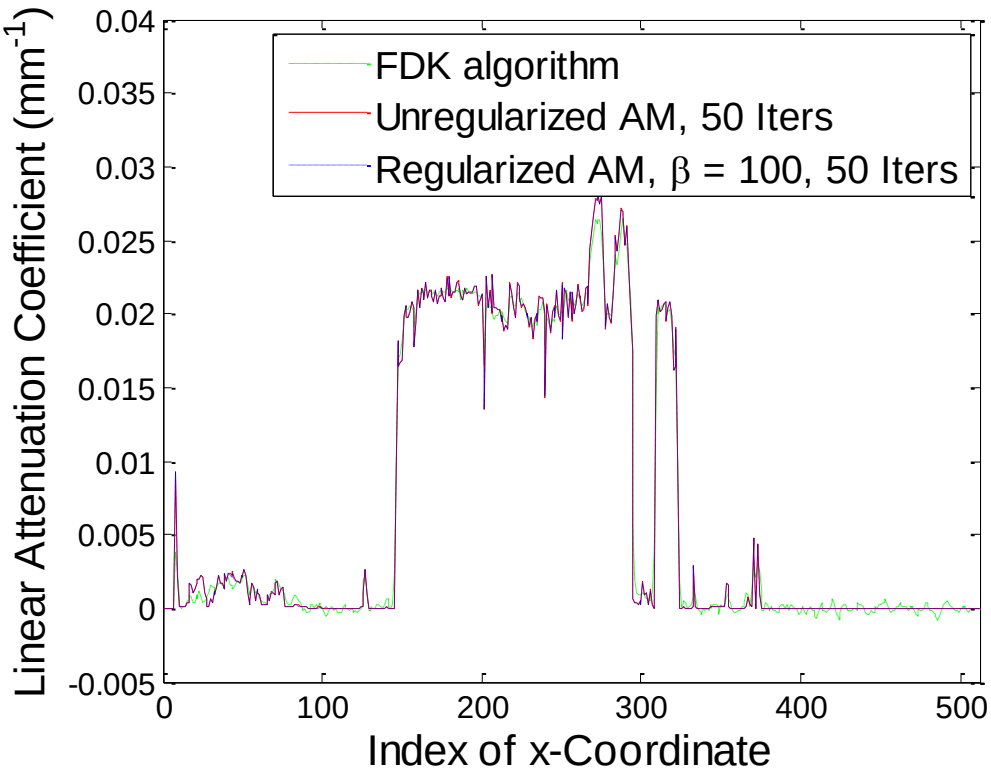
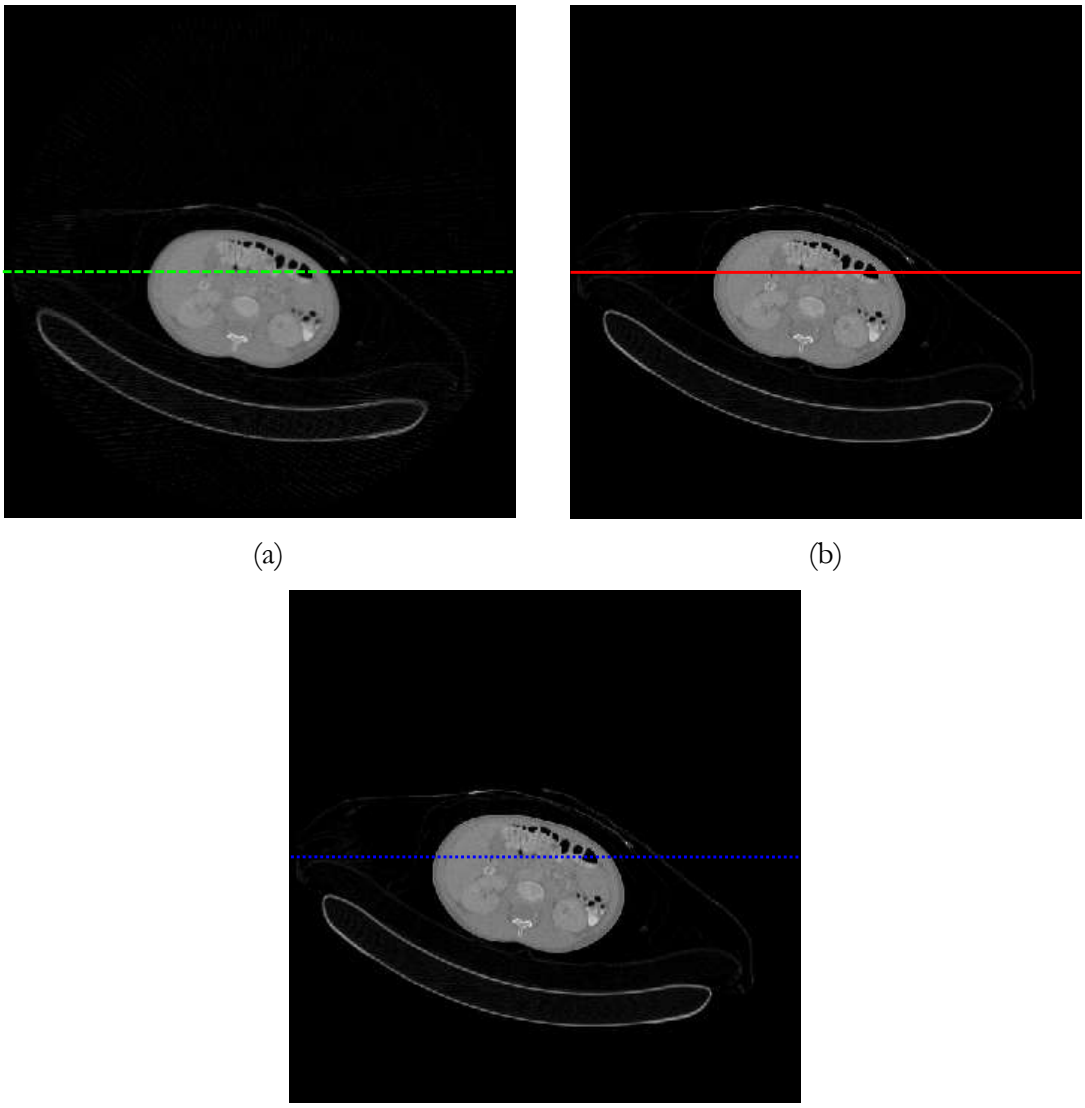


Figure 3.5 Profiles of the FDK, unregularized AM and regularized AM with $\beta = 100$ images shown in Figure 3.4. 1/2 of the views are used here. The profiles are plotted along the lines at row 256, as shown in Figure 3.4.

In this experiment, with every other view taken, all of the three algorithms still perform well. Figure 3.5 shows that the profiles are close to each other within the object. But in the region of air, FDK generates some fluctuations that are more obvious here than the ones in the full-data case. Next, we test a more extreme case which takes every fifth view.



(a) (b)

(c)

Figure 3.6 Image reconstruction from 1/5 of the views of data, all detectors. The 82nd axial slice is shown in all panels. (a) FDK algorithm, (b) unregularized AM algorithm, 50 iterations, and (c) regularized AM algorithm, penalty weight $\beta = 100$, 50 iterations. The display windows are $[0, 0.0418]$ mm^{-1}.

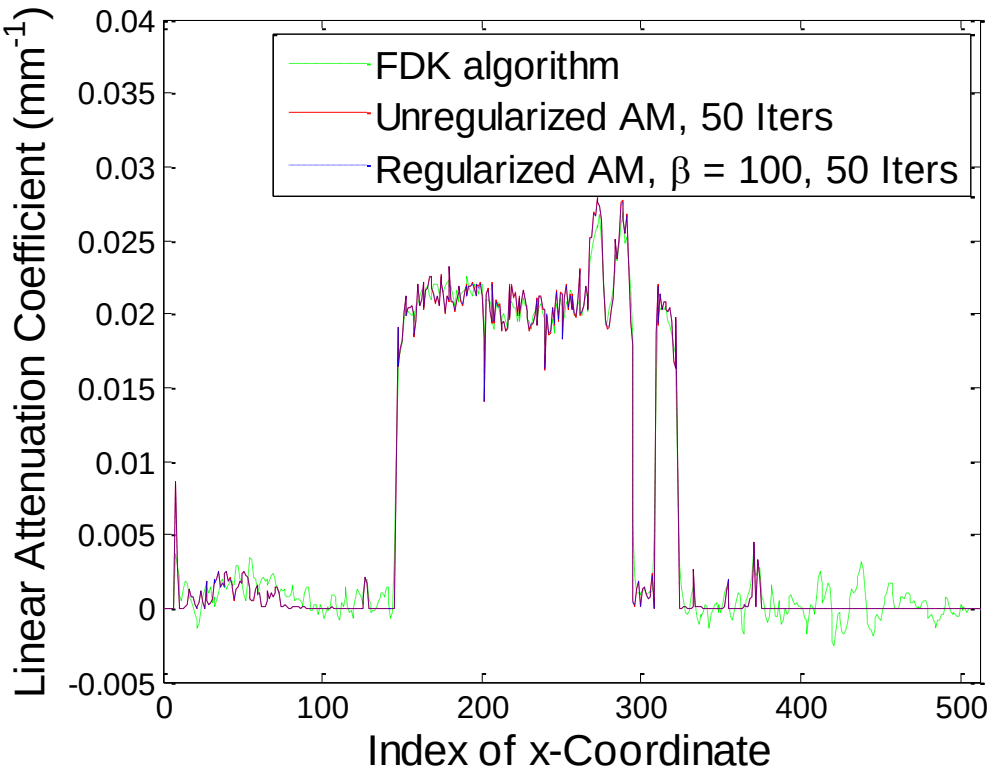
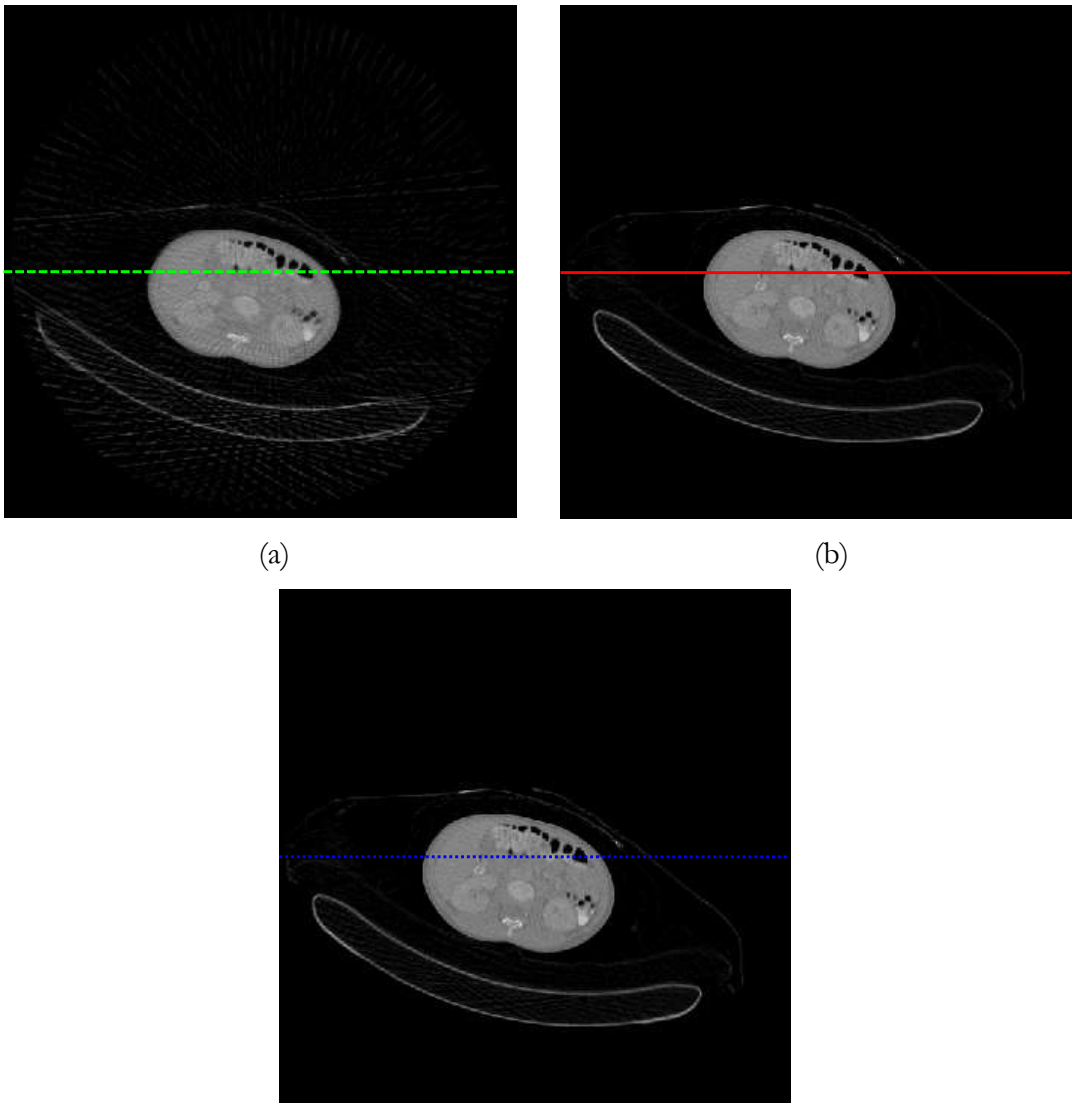


Figure 3.7 Profiles of the FDK, unregularized AM and regularized AM with $\beta = 100$ images shown in Figure 3.6. 1/5 views are used here. The profiles are plotted along the lines at row 256, as shown in Figure 3.6.

In this experiment, with every fifth view taken, we can find some obvious aliasing artifacts with FDK reconstruction, especially in the area of the bed. Unregularized AM and regularized AM perform well with this 1/5 views sampling. Figure 3.7 shows that in the region of air, FDK generates obvious fluctuations, which are stronger than the ones by taking every other view. Next, we test a more extreme case which takes every tenth view.



(a) (b)

(c)

Figure 3.8 Image reconstruction from 1/10 of the views of data, all detectors. The 82nd axial slice is shown in all panels. (a) FDK algorithm, (b) unregularized AM algorithm, 50 iterations, and (c) regularized AM algorithm, penalty weight $\beta = 100$, 50 iterations. The display windows are $[0, 0.0418]$ mm^{-1}.

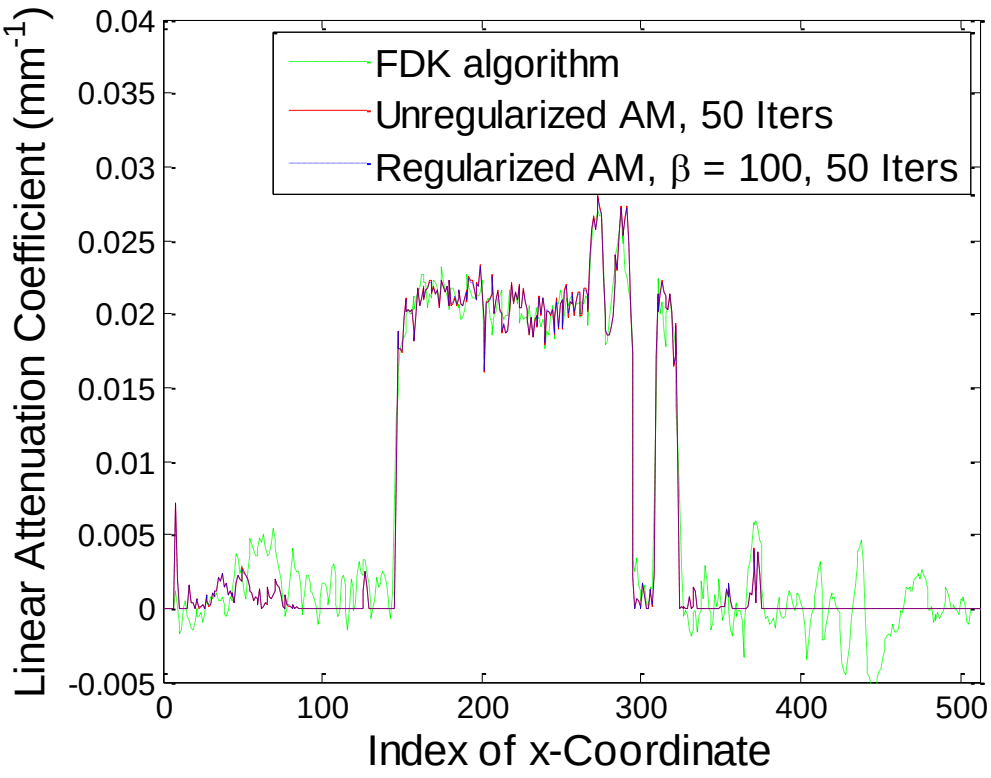
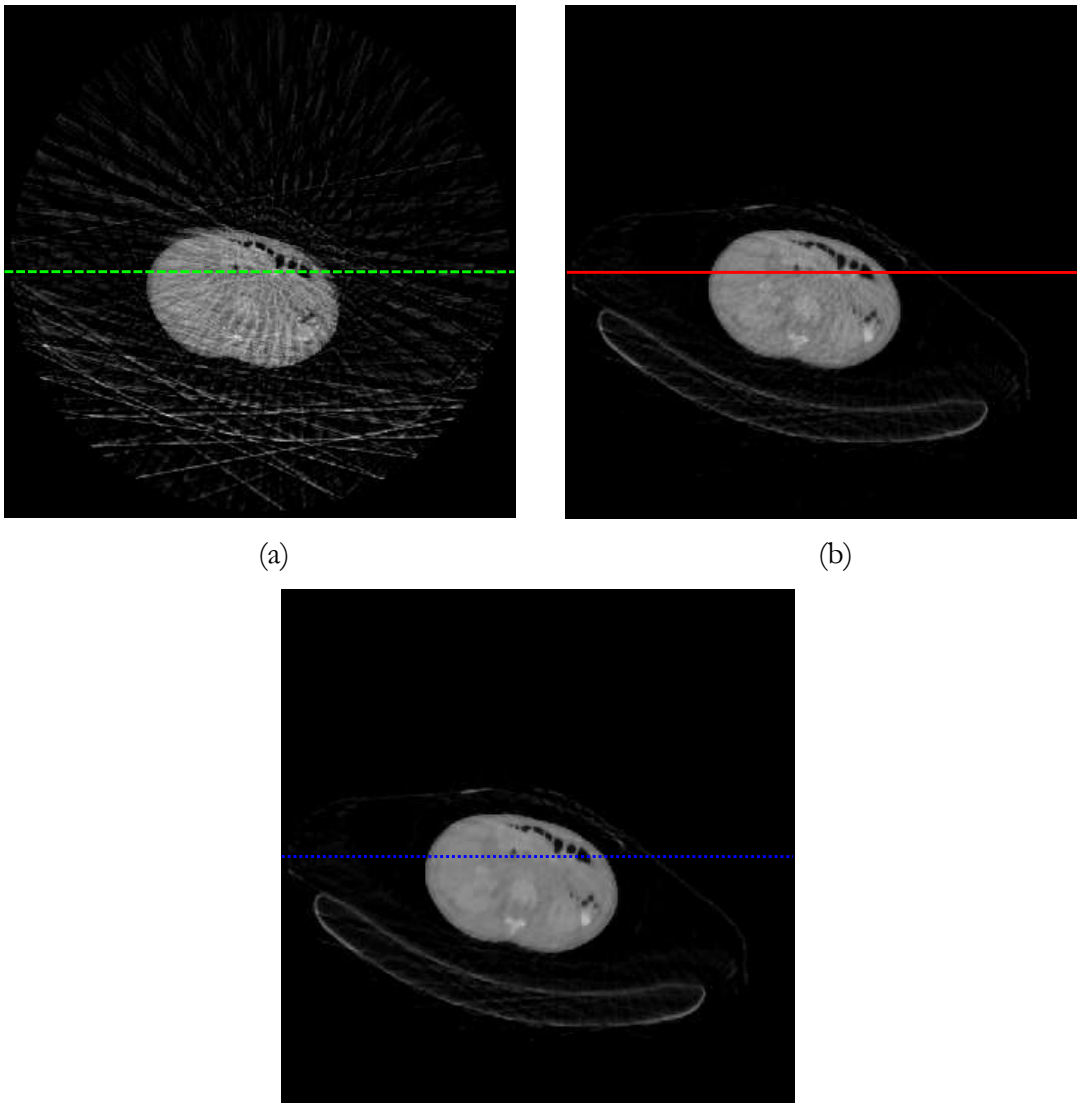


Figure 3.9 Profiles of the FDK, unregularized AM and regularized AM with $\beta = 100$ images shown in Figure 3.8. 1/10 views are used here. The profiles are plotted along the lines at row 256, as shown in Figure 3.8.

In this experiment, with every tenth view taken, more obvious aliasing artifacts from FDK reconstruction can be found, especially in the area of the bed. Unregularized AM and regularized AM start generate aliasing artifacts. Figure 3.9 shows that in the region of air, FDK generates more obvious fluctuations. Next, we test the most extreme case among our experiments for few-view sampling, which takes every twenty-ninth view.



(a) (b)

(c)

Figure 3.10 Image reconstruction from 1/29 of the views of data, all detectors. The 82nd axial slice is shown in all panels. (a) FDK algorithm, (b) unregularized AM algorithm, 50 iterations, and (c) regularized AM algorithm, penalty weight $\beta = 3200$, 50 iterations. The display windows are $[0, 0.0418]$ mm^{-1}.

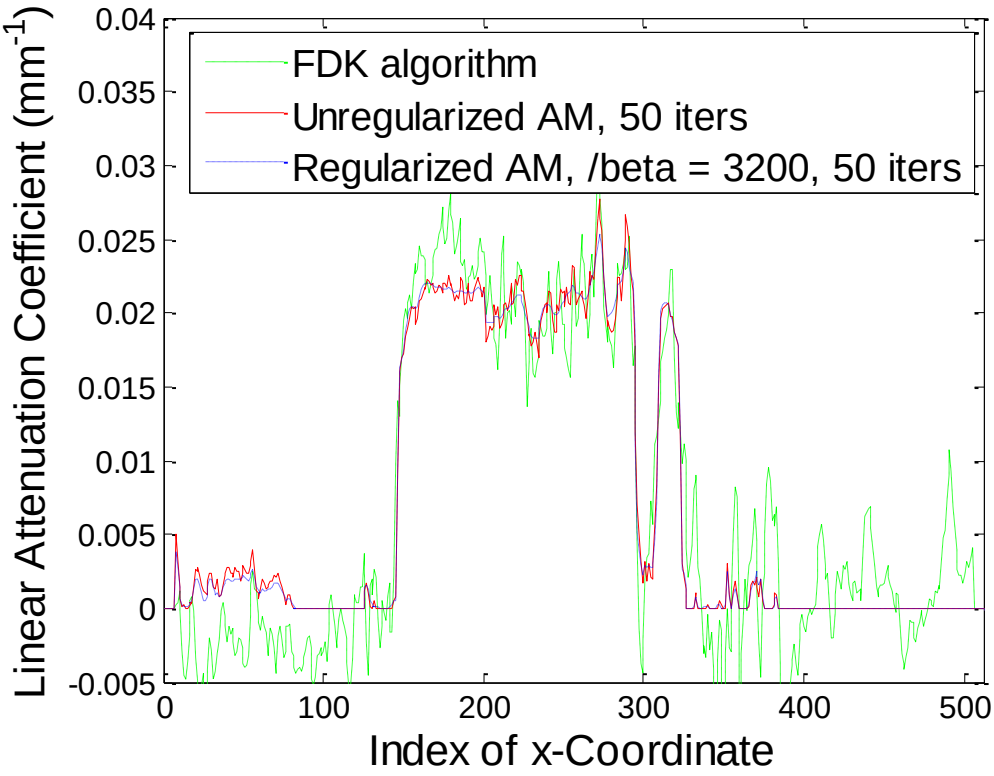
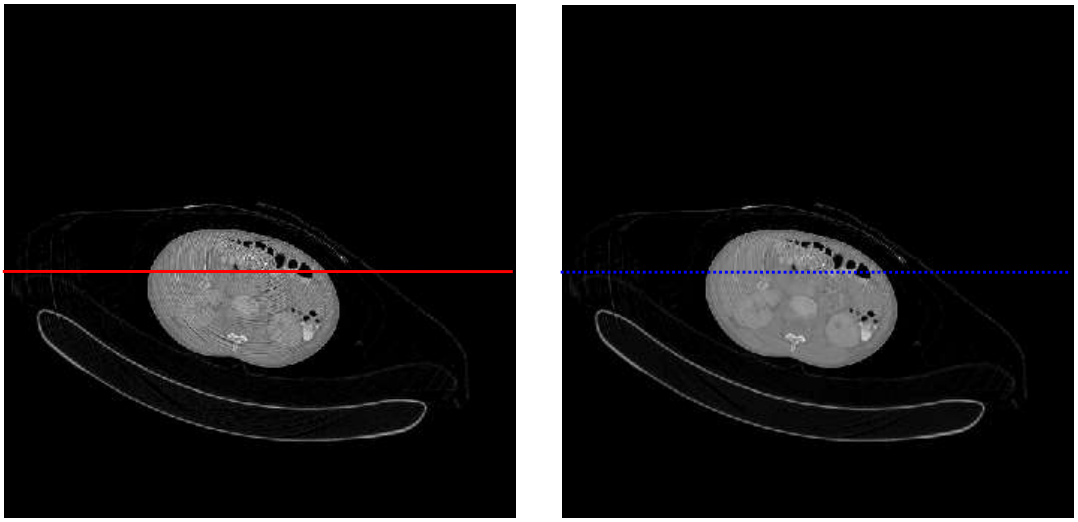


Figure 3.11 Profiles of the FDK, unregularized AM and regularized AM with $\beta = 100$ images shown in Figure 3.10. 1/29 views are used here. The profiles are plotted along the lines at row 256, as shown in Figure 3.10.

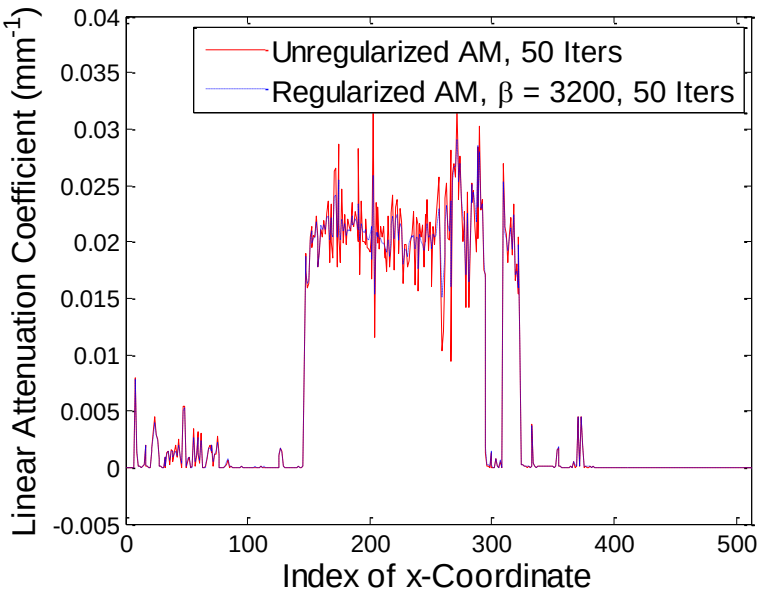
In this experiment, with every twenty-ninth view taken, severe aliasing artifacts from FDK reconstruction can be found and make the image quality extremely bad. Details are violated, which makes the image diagnostically useless. Unregularized AM generates obvious aliasing artifacts. To suppress those artifacts, $\beta = 3200$ is used. The aliasing artifacts are weakened to some extent, especially inside the object. However, the artifacts still exist and details are blurring. Figure 3.11 (c) shows that FDK generates insensible fluctuations and the fluctuations within the profile of unregularized AM are smoothed with regularization.

3.3.3 Subsampling by Detectors

In this section, we choose a subset of detectors to use in the reconstructions and discard the remaining detectors. As mentioned in section 3.3.1, we have two kinds of sampling strategies with subsampling by detectors: uniform pattern sampling and random pattern sampling. As for uniform pattern sampling, we randomly select a subset of detectors for the first view and use the same sampling pattern for all the other views. As for random pattern sampling, we randomly select a subset of detectors for each view independently. Specifically, the random selection is done by Poisson sampling, which selects detectors based on independent and identically distributed (i.i.d.) Bernoulli trials. Therefore, for the following experiments, the proportion of detectors being used is approximated. Starting with 1/2 data case, we progressively discard more and more detectors and plot the profile through the central cutting line to evaluate the performance.



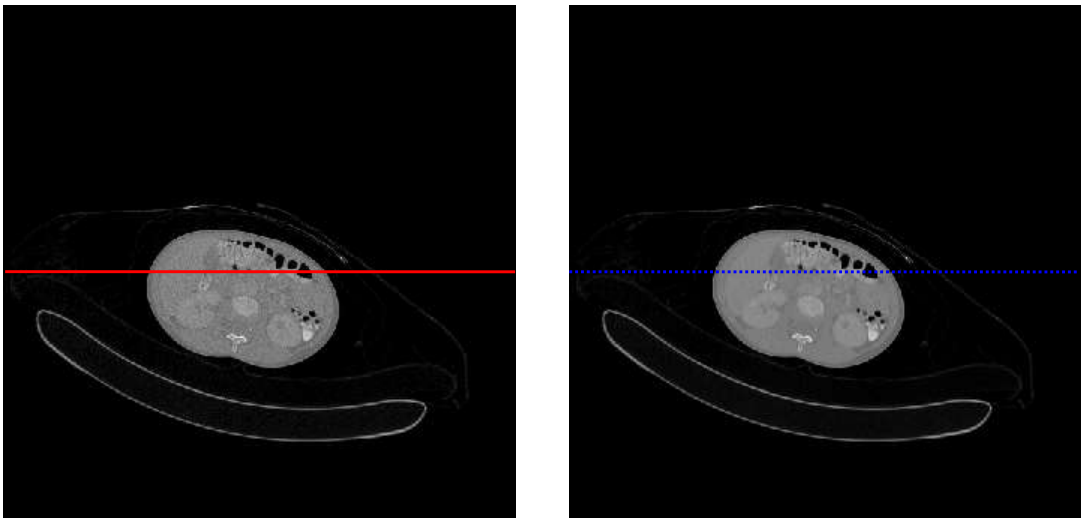
(a) (b)



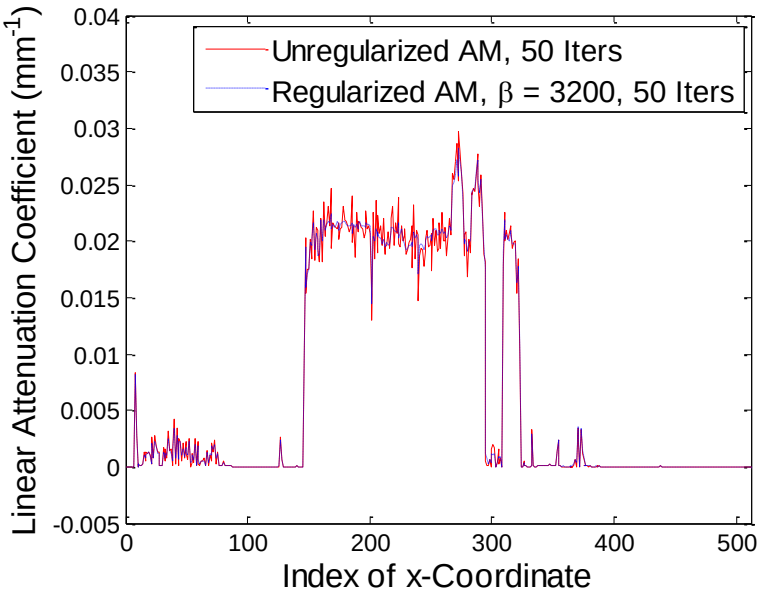
(c)

Figure 3.12 Image reconstruction from data collected from approximately 1/2 of the detectors, uniform pattern sampling, all views. The 82nd axial slice is shown in all panels. (a) unregularized AM algorithm, 50 iterations, (b) regularized AM algorithm, penalty weight $\beta = 3200$, 50 iterations, and (c) profiles of unregularized AM and regularized AM with $\beta = 3200$ from row 256. The display windows are $[0, 0.0418]$ mm^{-1}.

In this experiment, with 1/2 detectors uniform-pattern sampling, spiral artifacts appear when using unregularized AM. Regularized AM suppresses the spiral artifacts to some extent with $\beta = 3200$. Figure 3.12 (c) shows that the fluctuations within the profile of unregularized AM are smoothed with regularization. Next, we still use approximately 1/2 of the detectors, but with random pattern sampling.



(a) (b)

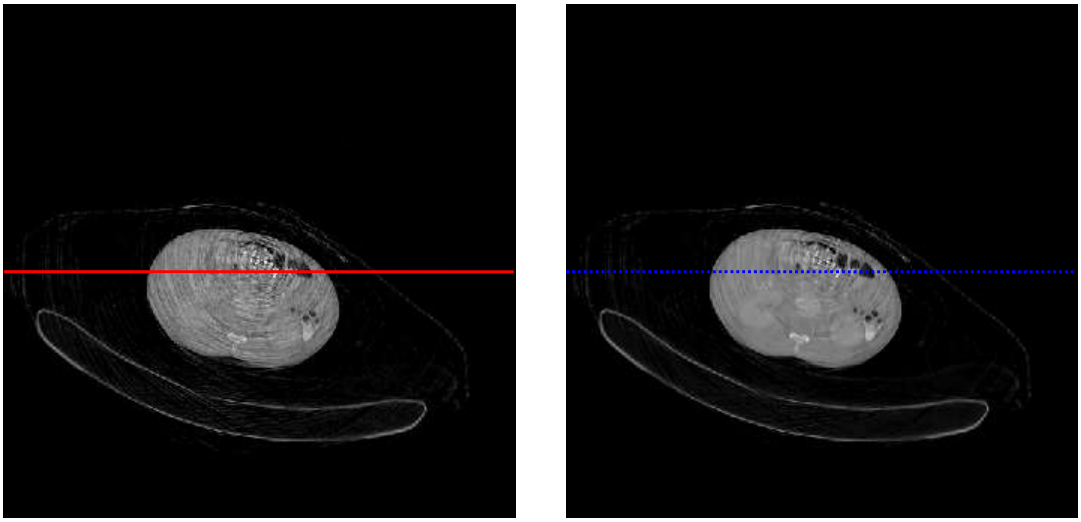


(c)

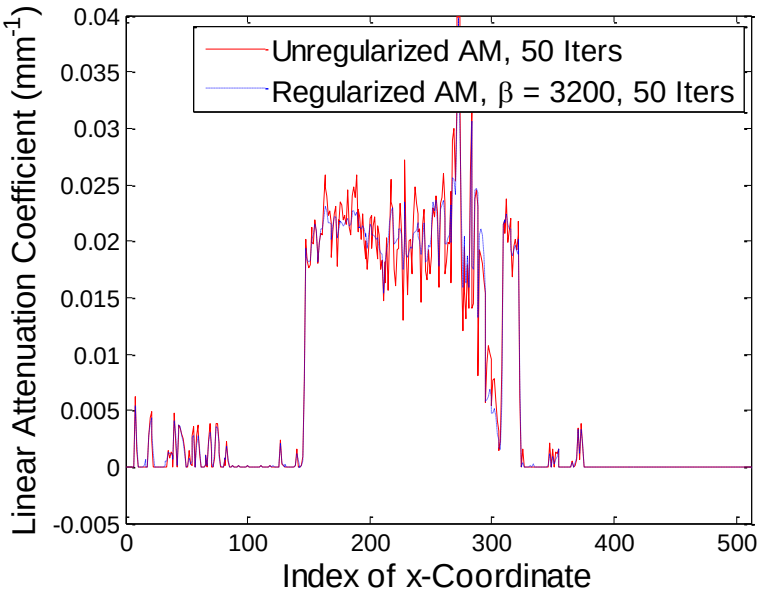
Figure 3.13 Image reconstruction from data collected from approximately 1/2 of the detectors, random pattern sampling, all views. The 82nd axial slice is shown in all panels. (a) unregularized AM algorithm, 50 iterations, (b) regularized AM algorithm, penalty weight $\beta = 3200$, 50 iterations, and (c) profiles of unregularized AM and regularized AM with $\beta = 3200$ from row 256. The display windows are $[0, 0.0418]$ mm^{-1}.

In this experiment, with 1/2 detectors random-pattern sampling, salt and pepper noise, instead of spiral artifacts, appear when using unregularized AM. Regularized AM suppresses the noise well with $\beta = 3200$. Figure 3.13 (c) shows that the fluctuations within the profile of unregularized AM are

smoothed well with regularization. Next, approximately 1/10 of the detectors are taken with uniform pattern sampling.



(a) (b)

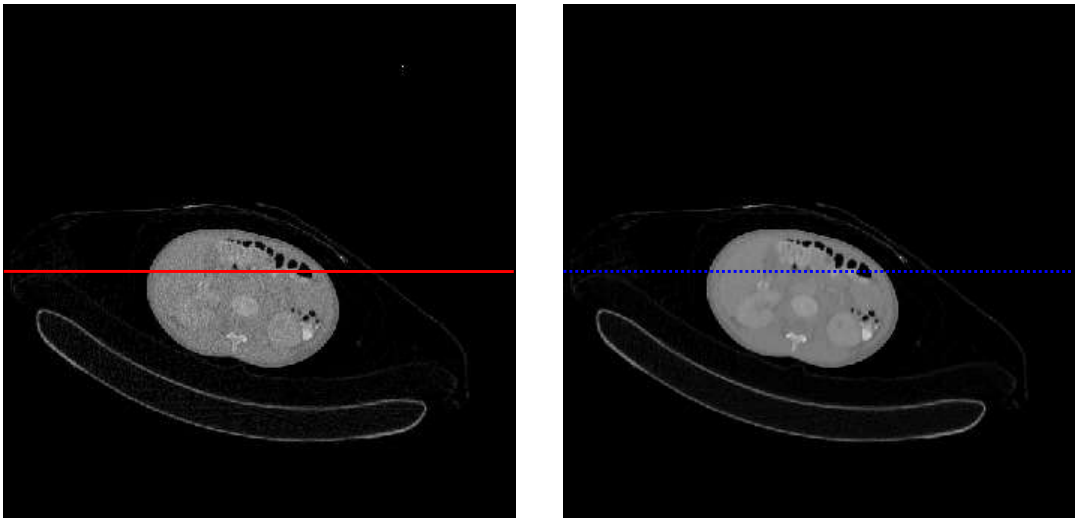


(c)

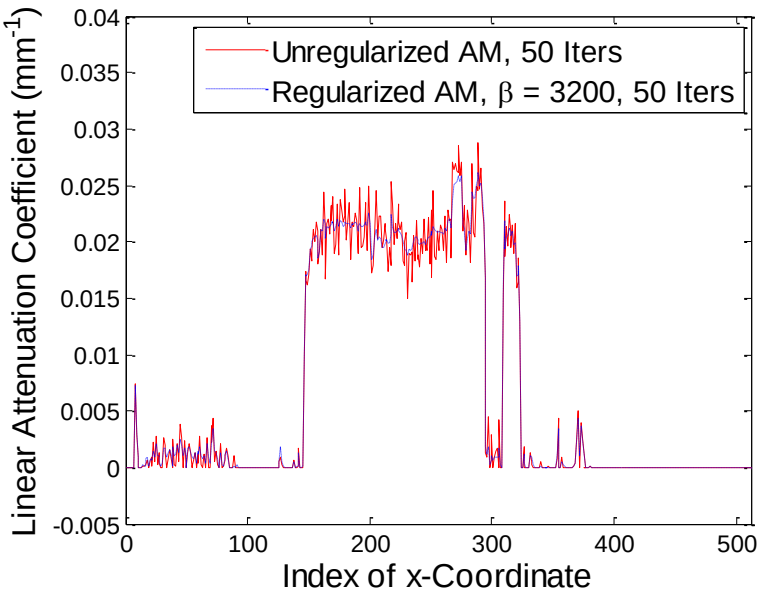
Figure 3.14 Image reconstruction from data collected from approximately 1/10 of the detectors, uniform pattern sampling, all views. The 82nd axial slice is shown in all panels. (a) unregularized AM algorithm, 50 iterations, (b) regularized AM algorithm, penalty weight $\beta = 3200$, 50 iterations, and (c) profiles of

unregularized AM and regularized AM with $\beta = 3200$ from row 256. Newton's method with a trust region was used. The display windows are $[0, 0.0418]$ mm^{-1}.

In this experiment, with 1/10 detectors uniform-pattern sampling, spiral artifacts get much stronger when using unregularized AM. Regularized AM suppresses the spiral artifacts to a great extent with $\beta = 3200$, but the artifacts are still obvious. Figure 3.14 (c) shows that the fluctuations within the profile of unregularized AM are smoothed with regularization. Next, we still use approximately 1/10 of the detectors, but with random pattern sampling.



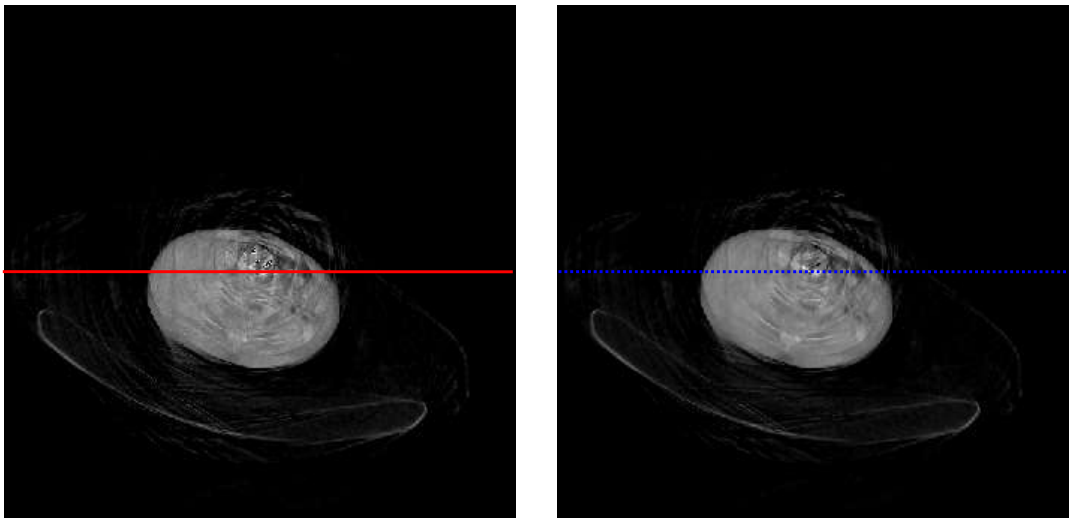
(a) (b)



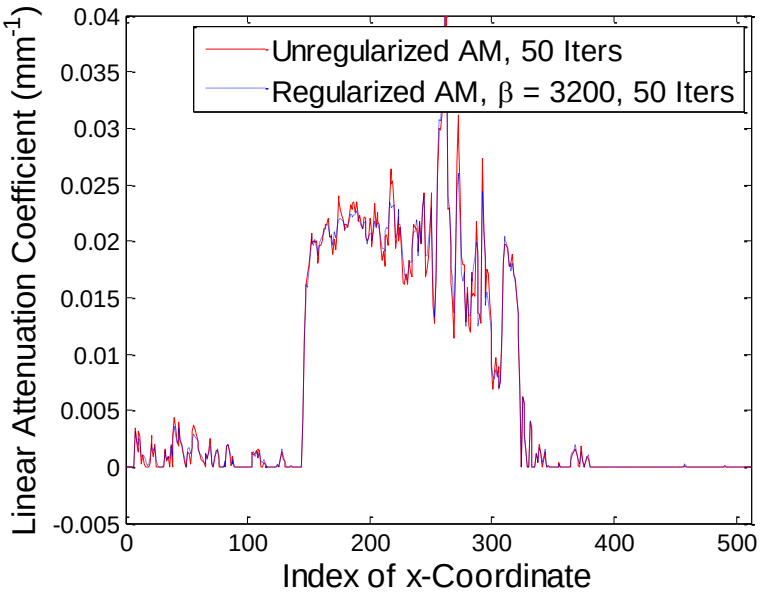
(c)

Figure 3.15 Image reconstruction from data collected from approximately 1/10 of the detectors, random pattern sampling, all views. The 82nd axial slice is shown in all panels. (a) unregularized AM algorithm, 50 iterations, (b) regularized AM algorithm, penalty weight $\beta = 3200$, 50 iterations, and (c) profiles of unregularized AM and regularized AM with $\beta = 3200$ from row 256. Newton's method with a trust region was used. The display windows are $[0, 0.0418]$ mm-1.

In this experiment, with 1/10 detectors random pattern sampling, salt and pepper noise gets stronger compared to the image using 1/2 detectors random pattern sampling with unregularized AM. Regularized AM suppresses the noise well with $\beta = 3200$. Figure 3.13 (c) shows that the fluctuations within the profile of unregularized AM are smoothed well with regularization. Next, approximately 1/29 of the detectors are taken with uniform pattern sampling.



(a) (b)

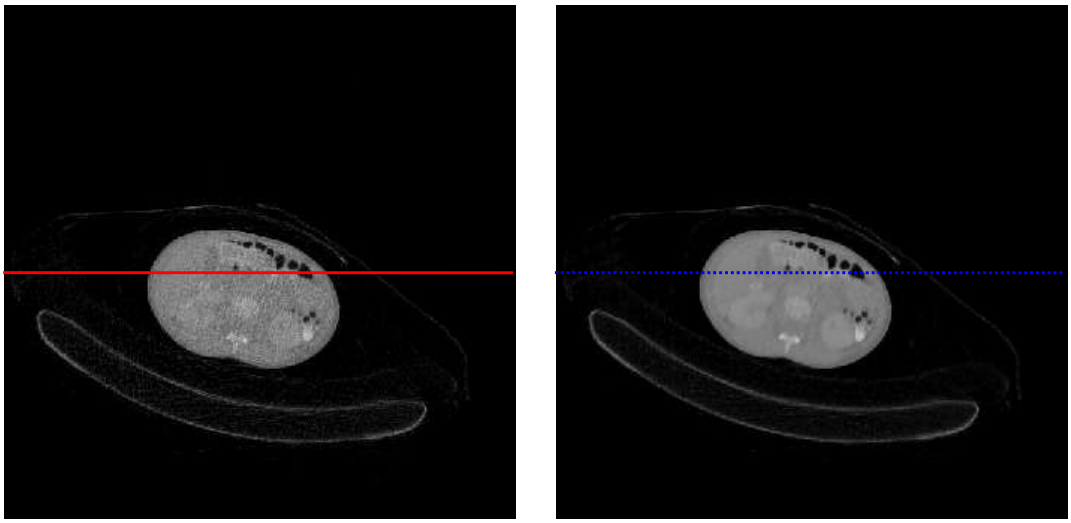


(c)

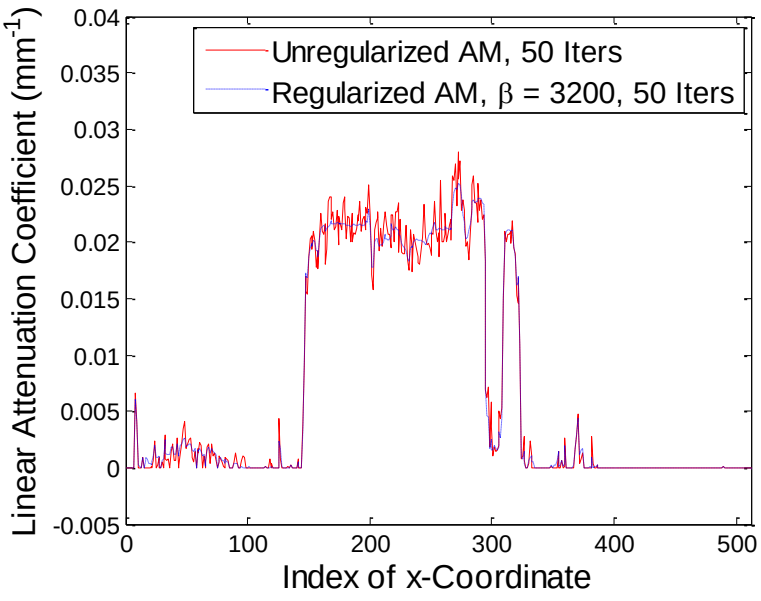
Figure 3.16 Image reconstruction from data collected from approximately 1/29 of the detectors, uniform pattern sampling, all views. The 82nd axial slice is shown in all panels. (a) unregularized AM algorithm, 50 iterations, (b) regularized AM algorithm, penalty weight $\beta = 3200$, 50 iterations, and (c) profiles of unregularized AM and regularized AM with $\beta = 3200$ from row 256. Newton's method with a trust region was used. The display windows are $[0, 0.0418]$ mm-1.

In this experiment, with 1/29 detectors uniform-pattern sampling, spiral artifacts are extremely strong when using unregularized AM. Even with regularization, the artifacts are still obvious and

details of the image are almost blurred out. Figure 3.14 (c) shows that the regularization does not help in this case. Next, we still use approximately 1/29 of the detectors, but with random pattern sampling.



(a) (b)



(c)

Figure 3.17 Image reconstruction from data collected from approximately 1/29 of the detectors, random pattern sampling, all views. The 82nd axial slice is shown in all panels. (a) unregularized AM algorithm, 50 iterations, (b) regularized AM algorithm, penalty weight $\beta = 3200$, 50 iterations, and (c) profiles of

unregularized AM and regularized AM with $\beta = 3200$ from row 256. Newton's method with a trust region was used. The display windows are $[0, 0.0418]$ mm^{-1}.

In this experiment, with 1/29 detectors random pattern sampling, salt and pepper noise gets even stronger compared to the image using 1/10 detectors random pattern sampling with unregularized AM. However, regularized AM suppresses the noise perfectly with $\beta = 3200$. The image quality in Figure 3.13 (b) is surprisingly good. Figure 3.13 (c) shows that the fluctuations within the profile of unregularized AM are smoothed well with regularization.

3.3.3 Conclusions of the Experiments

1. When using all 672x16 detectors and subsampling by views, FDK creates starburst-like aliasing artifacts and the image quality deteriorates when taking every tenth view. FDK algorithms are not suitable for image reconstruction from data with few views because they are derived from a continuous sampling model and therefore densely sampled projections are required to avoid aliasing.
2. AM avoids the starburst-like aliasing artifacts when taking every tenth view, but the image quality degrades when using every twenty-ninth view. The idea that optimization-based reconstruction with a discrete image model can be more effective than analytic-based reconstruction in dealing with few projections has been investigated by numerous authors [7][11][25]. I include my results here to illustrate this idea again.
3. When using all the views, reconstructing with uniform pattern subsampling of detectors using the AM algorithm creates a spiral artifact, especially at the central region of the image. The spiral artifact is more apparent when fewer detectors are used. The regularization method suppresses this artifact to some extent.
4. Reconstruction from random-pattern subsampling of the detectors using the AM algorithm avoids the spiral pattern but generates salt and pepper noise. However, with proper regularization, the salt and pepper noise is suppressed. When 1/29 of the detectors are used, random-pattern subsampling still performs well with proper regularization.

Chapter 4

Conclusions and future work

4.1 Conclusions

For regularized AM, Newton's method with the trust region modification is introduced to solve the decoupled image updating problem. The convergence of the method can be shown and in the sense of practical implementation, it reduces the chance of oscillations during the iterative updates. More importantly, this method can be readily implemented for a huge class of iterative algorithms and therefore has broad applications.

When taking partial views with all detectors used, the FDK algorithm generates apparent artifacts when the number of views drops below the number of views needed theoretically. This result matches our analysis well. With approximately 1/29 of the detectors, the image quality with uniform pattern sampling using either unregularized or regularized AM is extremely bad. However, when using random pattern sampling, although unregularized AM generates salt and pepper noise, the regularized AM suppresses the noise well and generates images with high quality, which is almost comparable to the one using regularized AM with full data. This result is interesting because, based on the idea from CS that random sampling with low average sampling rate might be able to capture most of the information in the signal, we have experimentally shown that random sampling in the whole detector space with only 1/29 of the data, we are able to restore most of the details of the object. Notice that with only 1/29 of the detectors being used, our optimization goal is to recover an image with size $512\times512\times164=42991616$ from measurements with size $\frac{672\times16\times13920}{29}=5160960$. Since the number of voxels to be estimated in the image is more than 8 times the number of measurements, the optimization problem is underdetermined. However, the results are better than we naively expected.

This subsampling topic is a very attractive research topic because of the numerous potential dose-reduction applications. To apply random pattern sampling to an X-ray CT system, one possible technique for reducing dose would be to add a random mask between the source and the patient to block some of the source-to-detector pairs and dynamically switch masks as the gantry rotates.

4.2 Future work

Newton's Method with the trust region modification has been shown to be effective in solving many computational problems. However, the current version of Newton's method is time-consuming. For a $512\times512\times164$ image volume, approximately 8 seconds are needed for each update. Due to the fact that the AM algorithm for transmission tomography needs lots of updates to converge, a modification is needed to increase the convergence speed of the Newton's method with the trust region.

Disregarding detectors randomly is equivalent to disregarding rows in the system matrix. This process increase the randomness of the system matrix and therefore the CS theories apply here. Random pattern sampling is probably the most "random" case with our current CT system design. However, more work can be done here to find the best choice of parameters for the regularized AM algorithm. Moreover, consider a hypothetical imaging system that is even more sparse so that, for example, it has exactly one nonzero element in each row of the system matrix. This means 7/8 of the voxels cannot be estimated. The point here is that from the perspective of computing an iterative algorithm, the more sparse the system matrix is, the easier or faster we can compute the forward and backward projections. On the other hand, from the perspective of designing an imaging system to reduce dose, or more generally speaking, designing a sampling protocol that requires fewer samples, the system matrix is expected to be less sparse, so that it is possible to take even fewer measurements to recover a large image. So, for future work, a challenge could be how to design a CT system with a less sparse system matrix based on CS theories.

Appendix A

This appendix shows the derivation of the regularized AM algorithm. It was shown in [6] and [22], but is included here as well for convenience.

First, the $\psi\left(c(x)-c(x')\right)$ terms of the penalty function is decoupled. Due to the convexity of ψ, we have

$$\psi\left(c(x)-c(x')\right) \le \frac{1}{2}\psi\left\{2[c(x)-\hat{c}(x))+(\hat{c}(x)-c(x')]\right\}+\frac{1}{2}\psi\left\{2(\hat{c}(x')-c(x'))+(\hat{c}(x)-\hat{c}(x')\right\},$$

$$=\frac{1}{2}\psi\left\{2c(x)-\hat{c}(x)-\hat{c}(x')]\right\}+\frac{1}{2}\psi\left\{2c(x')-\hat{c}(x)-\hat{c}(x')]\right\}.$$

Here, $\hat{c}(\cdot)$ refers to the current image estimate.

By plugging this surrogate into our objective function $R(c)$ and denoting the new objective function by $R_{new}(c)$, our goal is to solve

$$\frac{\partial I[d(y), y\in\mathcal{Y} \mid q(y:c), y\in\mathcal{Y}]}{\partial c(\mathrm{x})}+\beta\frac{\partial R_{new}(c)}{\partial c(\mathrm{x})}=0.$$

We denote two backprojections, data backprojection and mean backprojection, as follows.

$$b(\mathrm{x})=\sum\nolimits_{y\in\mathcal{Y}} h(y\mid x)d(y),$$

$$\hat{b}(\mathrm{x})=\sum\nolimits_{y\in\mathcal{Y}} h(y\mid x)q(y:\hat{c}(x)),$$

Following [1], we obtain the following equation.

$$b(\mathrm{x})-\hat{b}(\mathrm{x})\exp\left[Z(\mathrm{x})\left(\hat{c}(x)-c(\mathrm{x})\right)\right]+\beta\sum_{x'\in N_x}\omega_{x,x'}\frac{\partial\psi(t)}{\partial t}=0,$$

where $t=2c(\mathrm{x})-\hat{\mathrm{c}}(\mathrm{x})-\hat{\mathrm{c}}(\mathrm{x}').$

HECTARE uses Newton's method to solve this equation. The first derivative and second derivative are calculated for each Newton's step using

$$\frac{\partial I[d(y), y\in\mathcal{Y} \mid q(y:c), y\in\mathcal{Y}]}{\partial c(\mathrm{x})}=b(\mathrm{x})-\hat{b}(\mathrm{x})\exp\left[Z(\mathrm{x})\left(\hat{c}(x)-c(\mathrm{x})\right)\right],$$

$$\frac{\partial}{\partial c(\mathrm{x})}\left(\frac{\partial I[d(y), y \in \mathcal{Y} \mid q(y:c), y \in \mathcal{Y}]}{\partial c(\mathrm{x})}\right) = Z(\mathrm{x})\hat{b}(\mathrm{x})\exp\left[Z(\mathrm{x})\left(\hat{c}(x) - c(\mathrm{x})\right)\right],$$

$$\frac{\partial R_{new}(c)}{\partial c(\mathrm{x})} = \sum_{x' \in N_x} \omega_{x,x'} \frac{t}{1+|\delta t|},$$

$$\frac{\partial}{\partial c(\mathrm{x})}\left(\frac{\partial R_{new}(c)}{\partial c(\mathrm{x})}\right) = 2\sum_{x' \in N_x} \omega_{x,x'} \frac{1}{\left(1+|\delta t|\right)^2},$$

where $t = 2c(\mathrm{x}) - \hat{\mathrm{c}}(\mathrm{x}) - \hat{\mathrm{c}}(\mathrm{x}').$

Therefore, the overall updating steps in HECTARE for regularized AM can be written as

$$c^{(\mathrm{k}+1)}(\mathrm{x}) = \mathrm{c}^{(\mathrm{k})}(\mathrm{x}) - \frac{b(\mathrm{x}) - \hat{b}(\mathrm{x})\exp\left[Z(\mathrm{x})\left(\hat{c}(x) - \mathrm{c}^{(\mathrm{k})}(\mathrm{x})\right)\right] + \beta\sum_{x' \in N_x} \omega_{x,x'} \frac{t}{1+|\delta t|}}{Z(\mathrm{x})\hat{b}(\mathrm{x})\exp\left[Z(\mathrm{x})\left(\hat{c}(x) - \mathrm{c}^{(\mathrm{k})}(\mathrm{x})\right)\right] + 2\beta\sum_{x' \in N_x} \omega_{x,x'} \frac{1}{\left(1+|\delta t|\right)^2}},$$

where k is the Newton's updating iteration index and $t = 2c^{(\mathrm{k})}(\mathrm{x}) - \hat{\mathrm{c}}(\mathrm{x}) - \hat{\mathrm{c}}(\mathrm{x}').$

References

[1] J. Hsieh, *Computed Tomography: principles, design, artifacts, and recent advances*, *SPIE*, *Bellingham*, *WA*, 2009.

[2] D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, and A. A. Martinez, *Flat-panel cone-beam computed tomography for image-guided radiation therapy*, *Int. J. Radiation Oncology Biol. Phys.* vol. 53, no. 5, pp. 1337-1349, 2002.

[3] X. Han, J. Bian, E. L. Ritman, E. Y. Sodky, and X. Pan, *Optimization-based reconstruction of sparse images from few-view projections*, *Phys. Med. Biol.* , vol. 57, pp. 5245-5273, 2012.

[4] X. Li and S. Luo, *A compressed sensing-based iterative algorithm for CT reconstruction and its possible application to phase contrast imaging*, *BioMedical Engineering Online*, 2011.

[5] E. J. Candès and M. B. Wakin, *An introduction to compressive sampling*, *IEEE Signal Processing Magazine*, vol. 21, March 2008.

[6] D. L. Donoho, *Compressed sensing, IEEE Transactions on Information Theory*, vol. 52, no. 4, 2006.

[7] G. Chen, J. Tang, and S. Leng, *Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets*, Med. Phys. Vol. 35, 660–663, 2008.

[8] M. G. Lubner, P. J. Pickhardt, J. Tang, and G. Chen, *Reduced image noise at low-dose multidetector CT of the abdomen with prior image constrained compressed sensing algorithm*, *Radiology*, vol. 260(1): pp. 248-256, 2011.

[9] K. Choi, J. Wang, L. Zhu, T. Suh, S. Boyd, and L. Xing, *Compressed sensing based cone-beam computed tomography reconstruction with a first-order method*, *Med. Phys.* vol. 37, No. 9: pp. 5113-5125, 2010.

[10] R. G. Baraniuk, *Compressive sensing*, *IEEE Signal Processing Mag.*, vol. 24, No. 4, pp. 118-120, 124, 2007.

[11] E. Y. Sidky, C. Kao, and X. Pan, *Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT*, *Journal of X-Ray Science and Technology*, Vol. 14, 119-139, 2006.

[12] D. B Keesing, *Development and implementation of fully 3D statistical image reconstruction algorithms for helical CT and half-ring PET insert system*, PhD dissertation, Washington Univ. in St. Louis, 2009.

[13] J. L. Prince and J. M. Links, *Medical Imaging Signals and Systems*, *Pearson Prentice Hall Bioengineering*, *Upper Saddle River, NJ,* 2004. Print.

[14] J. A. O'Sullivan and J. Benac, *Alternating minimization algorithms for transmission tomography*, *IEEE Trans. Med. Imag.*, vol. 26, pp. 283-297, Mar 2007.

[15] D. L. Synder, J. A. O'Sullivan, R. J. Murphy, D. G. Politte, B. R. Whiting and J. F. Williamson, *Image reconstruction for transmission tomography when projection data are incomplete*, *Phys. Med. Biol.* , vol. 51, pp. 5603-5619, Oct 2006.

[16] D. B. Keesing, *Description and usage guide for HECTARE v. 1.0, Helical CT Advanced Reconstruction Engine,* 2011.

[17] A. Katsevich, *An improved exact filtered backprojection algorithm for spiral computed tomography*, *Advances in Applied Mathematics*, vol. 32, no. 4, pp. 681-697, 2003.

[18] K. Stierstorfer, A. Rauscher, J. Boese, H. Bruder, S. Schaller, and T. Flohr, *Weighted FBP – a simple approximate 3D FBP algorithm for multislice spiral CT with good dose usage for arbitrary pitch*, *Phys. Med. Biol.*, vol. 49, pp. 2209-2218, 2004.

[19] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, *Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction, IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 166-175, 1997.

[20] D. G. Politte, S. Yan, J. A. O'Sullivan, D. L. Snyder, and B. R. Whiting, *Implementation of alternating minimization algorithms for fully 3D CT imaging, in Proc. SPIE: Computational Imaging III* (C. A. Bouman and E. L. Miller, eds.), vol. 5674, pp. 362-373, 2005.

[21] D. C. Sorensen, *Newton's method with a model trust region modification, Society for Industrial and Applied Mathematics,* vol. 19, No. 2, pp. 409-426, 1982.

[22] S. Ahn, J. A. Fessler, D. Blatt, and A. O. Hero, *Convergent incremental optimization transfer algorithms: application to tomography*, *IEEE Transactions on Medical Imaging,* vol. 25, No. 25, pp. 283-296, 2006.

[23] H. M. Hudson and R. S. Larkin, *Accelerated image reconstruction using ordered subsets of projection data, IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601-609, 1994.

[24] H. Erdogan and J. A. Fessler, *Ordered subsets algorithms for transmission tomography, Phys. Med. Biol.*, pp. 2835-2851, 1999.

[25] H. M. Hudson and R. S. Larkin, *Accelerated image reconstruction using ordered subsets of projection data, IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601-609, 1994.

[26] B. Hasegawa, *Physics of Medical X-Ray Imaging*, 2nd ed, 1987. Print.

[27] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. B. Wakin, *A simple proof of the restricted isometry property for random matrices*, *Constructive Approximation,* 2007.

[28] E. Y. Sidky, C. Kao, and X. Pan, *Effect of the data constraint on few-view fan-beam CT image reconstruction by TV minimization, IEEE Nuclear Science Symposium Conference*, 2006.